HARVESTING NOVEL BIOCATALYSTS FROM THE METAGENOME

This work was financially supported by the Dutch Technology and Science Foundation (STW) under project GCH.4919.

RIJKSUNIVERSITEIT GRONINGEN

HARVESTING NOVEL BIOCATALYSTS FROM THE METAGENOME

Proefschrift

ter verkrijging van het doctoraat in de
Wiskunde en Natuurwetenschappen
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. F. Zwarts,
in het openbaar te verdedigen op
vrijdag 9 juli 2004
om 14.45 uur
door

Esther Michèle Gabor

geboren op 26 juli 1973
te Mannheim, Duitsland

Promotor:	Prof. dr. D.B. Janssen
Beoordelingscommissie:	Prof. dr. L. Dijkhuizen
	Prof. dr. W.J. Quax
	Prof. dr. D. van Elsas

ISBN: 90-367-2056-7

CONTENTS

1

General introduction

The number of biotechnological production processes in the fine chemicals/pharmaceuticals sector has been steadily rising during the past decades (Schmid et al., 2001). A prominent example is the production of β-lactam antibiotics, which cover about 65 % of the world market for anti-infectives. Already since the 1940's, penicillin G, penicillin V, and cephalosporin C are biotechnologically produced by fungal fermentation. Nowadays, biocatalytic steps are also increasingly employed in the conversion of these natural antibiotics to the more effective semi-synthetic antibiotics, which was originally carried out via complex chemical routes that involve plenty of reaction steps and the production of large amounts of waste (Bruggink and Roy, 2001; Elander, 2003). Finding a suitable biocatalyst takes a key position in the development of a successful biocatalytic process. This task requires a sound searching strategy, starting from the definition of the actual search criteria and including the selection of appropriate biological source materials and screening methods. Possibly, the recovered best candidate for the envisaged application still needs to be optimized. In this introductory chapter, different aspects of biocatalyst discovery will be discussed. To limit the scope, we will solely focus on isolated enzymes and will not consider biotransformations that rely on whole cells. Furthermore, the biocatalytic importance of penicillin acylases, the group of enzymes mainly targeted in the presented research, will be shortly introduced.

1. The quest for the Holy Grail: the ideal biocatalyst

A recent OECD study on "The application of biotechnology to industrial sustainability" (2001) revealed that commercial processes applying biotechnology are in general more environmentally friendly than the purely chemical ones they are replacing. This is attributed to the mild conditions (low temperature and pressure) at which enzyme-catalyzed reactions can be run and the high specificity of biocatalysts, preventing the accumulation of large amounts of side-products. The use of potentially hazardous organic solvents is also reduced since enzymatic reactions can generally be carried out in aqueous medium. Still, efforts are being made to develop enzymes that also function in organic solvents, since those media make it possible to use hydrolytic enzymes for coupling reactions and often allow higher substrate or product concentrations. Another advantage of biological catalysts is the fact that they are obtained from renewable sources and are biodegradable, while the recycling or deposition of transition metal catalysts can constitute a major environmental problem.

In order to be competitive with traditional routes, however, biocatalytic conversions do not only need to increase the sustainability of a given process, but they must also deliver a clear cost advantage, e.g. by reducing the material and energy input or the production of waste. Whether or not a specific biotransformation step is economically feasible is also determined by the properties of the employed enzyme such as its stability, kinetic properties, enantioselectivity, the need for cofactors, and by the production and operating costs of the biocatalyst. Due to this, reaction conditions

are usually adjusted to the requirements of the chosen enzyme, rather than to the thermodynamic reaction optimum or the nature of substrates and products, leading to a compromised production process. With an increasing number of enzymatic activities being available, however, this paradigm of process design may more and more be reversed by selecting a biocatalyst that perfectly fits to the constraints of a given conversion (Burton et al., 2002) (Fig.1).

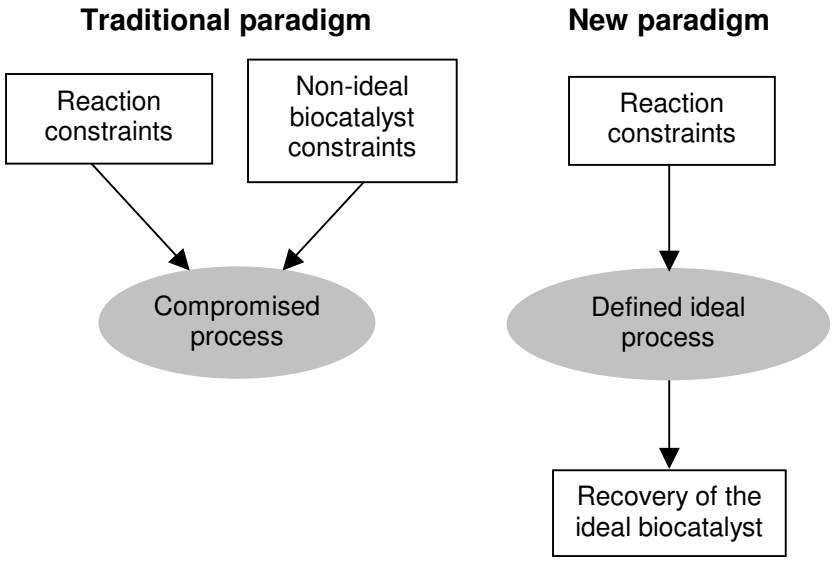


Figure 1. The paradigm shift (Burton et al., 2002). Nowadays, processes are usually designed to meet the requirements of the chosen biocatalyst. In the future, however, new enzyme discovery and improvement technologies may allow to find the ideal biocatalysts for any process that is defined on basis of the reaction constraints, e.g. thermodynamic equilibrium, or solubility and stability of educts and products.

But how to obtain such an "ideal" enzyme? In principle, the demand for a new enzyme with specific properties can be satisfied in two ways: the improvement of an already available enzyme by protein engineering or the recovery of a completely new protein from the natural gene pool. While the first approach explores sequence diversity in a relatively narrow and targeted fashion, using very specified genetic information as a starting material, the latter strategy bears a higher chance of accessing truly novel activities, since it is solely limited by the natural diversity of activities and the tools to access them. Unfortunately, the natural evolution of enzymes is not steered at providing efficient biocatalysts but at enabling organisms to compete in their natural habitats. One can thus anticipate that the properties of an enzyme found in nature normally do not fit the strict constraints of a given biocatalytic process. Therefore, the two above-mentioned ways of obtaining new enzymes should not be regarded as alternatives but as complementary methods for obtaining a desired biocatalyst.

2. Discovery of microbial enzymes

2.1. Culture screening

Until recently, pure cultures of microorganisms, obtained either from strain collections or enriched from natural habitats, have been the major starting point for enzyme discovery. Screening of pure cultures for their ability to perform a certain reaction and subsequent cloning of the responsible gene(s) is a productive way to recover new biocatalysts and, to date, about 11,000 fungal, 13,000 bacterial, and 440 archaeal species have been described [National Center for Biotechnology Information (NCBI), taxonomy browser statistics, http://www.ncbi.nlm.nih.gov/Taxonomy/txstat.cgi]. These isolates are though believed to constitute only a minor fraction of the microbial diversity present in nature, which may comprise millions of species (Torsvik et al., 2002). This notion is strongly supported by data from diverse habitats that were obtained during the last two decades, using molecular biological techniques (e.g. Amann and Kühl, 1998; Bintrim et al., 1997; Hugenholtz et al., 1998; Torsvik et al., 1990). The collective genomes present in only a single unperturbed organic soil sample, for instance, were found to equal at least 6,000-10,000 *E. coli* genomes in size, using DNA reassociation experiments (Torsvik et al., 1998).

Since in many habitats, more than 99 % of the bacterial cells observed under the microscope is uncultivable (Amann et al., 1995), microbial populations have remained largely unexplored by classical microbiological techniques (Table 1). The expression "uncultivable" is though somewhat misleading, as it does not refer to a well-defined intrinsic property of an organism, but describes the outcome of an experiment, namely the attempt to enrich it. However, defined growth conditions that are close enough to those prevailing in the natural (micro)habitats of microorganisms are often difficult to establish.

Table 1. Cultivability of bacteria in different habitats (adapted from Amann et al., 1995)

Habitat	Cultivability [%][a]
Seawater	0.001-0.1
Freshwater	0.25
Mesotrophic lake	0.1-1
Unpolluted estuarine waters	0.1-3
Activated sludge	1-15
Sediments	0.25
Soil	0.3

[a] Cultivability was determined as percentage of cultivable cells (cfu) in comparison with total cell counts.

In a recent report, Zengler et al. (2002) tackled this problem by incubating single cells encapsulated in agar-microdroplets with nutrient extracts prepared from their original habitats (seawater and soil) in a continuous system. After formation of

microcolonies was detected by flow cytometry, pure cultures could be established in rich medium. By this method, organisms were proliferated that fell into taxonomic groups that did not comprise any previously cultivated representatives. However, even when mimicking natural habitats quite well, many bacteria are expected to remain recalcitrant to isolation in pure culture due to their strict interdependence with other organisms (Kaeberlein et al., 2002) or because they are (temporarily) in a non-cultivable state, e.g. as a result of exposure to salt water, freshwater, low temperatures, or due to prolonged nutrient starvation conditions, as has been described for some bacterial organisms (Oliver et al., 1991; Roszak et al., 1984). For these reasons, enzyme discovery strategies that rely on growth probably still fail to exploit the majority of the microbial gene pool.

2.2. Metagenome screening

Culture-independent methods for biocatalyst discovery do not require the proliferation of microorganisms, but solely rely on the genetic information stored in the collective DNA of all microorganisms present in an environmental sample, the so-called metagenome (Handelsman et al., 1998). As the DNA isolated from an environmental sample theoretically encodes the enzymes of all indigenous microorganisms, an enormous variety of different biocatalysts can in principle be recovered from a single DNA sample. An example for the impressive diversity of metagenome-encoded enzymes was provided by Diversa Corporation (San Diego). By fragmenting total DNA from an alkaline desert sample, cloning it into an expression vector, and screening for esterase/lipase activity in an easily cultivable host strain, 120 new enzymes were discovered, falling into 21 protein families (Miller, 2000).

During the past five years, cloning of the metagenome has become the most popular tool for cultivation-independent enzyme discovery and led to the recovery of a range of new biocatalysts by academic and commercial institutions (Table 2). Many of these enzymes belong to novel biochemical pathways and, to our knowledge, none of the biocatalysts ever recovered from the metagenome was identical to an already known one. In almost all studies, *E. coli* was employed as expression host, but an example of cloning in a broad host range vector and expression in *Streptomyces lividans* has also emerged in the literature (Courtois et al., 2003). Vector systems used for the cloning of environmental DNA range from small-insert cloning vectors such as plasmids or phage vectors (up to 15 kb inserts) to bacterial artificial chromosomes (BACs) that can harbor as much as 100-kb fragments. While BAC vectors are most often applied when activities are targeted that may be dependent on the expression of large gene clusters (e.g. metabolite formation), small-insert libraries are usually prepared for the screening of single genes or small operons. The latter is often preferred because of the more convenient experimental procedures that can be used when working with small DNA fragments, concerning e.g. DNA purification and cloning. However, the smaller the cloned fragments are, the larger the resulting gene banks need to be for a comparable coverage of the genetic information, which requires more laborious screening procedures. For example, one amylase-expressing clone could be isolated per 450 clones screened when using a BAC vector, while 3,000 to

Table 2. Enzymes and pathways discovered by random cloning of environmental DNA and expression screening

Activity	Analyzed habitat	Vector (average insert size, kb)	Gene bank size (positive hits)	Reference
Alcohol oxidoreductase	Soil / sediment enrichment	Plasmid (4)	1,200,000 (16)	Knietsch et al., 2002
Amidase	Soil / sediment (+ enrichment)	Plasmid (5)	193,000 (6)	***Chapter 4***
Amylase	Soil	Plasmid (5)	80,000 (1)	***Chapter 2***
	Deep-sea enrichment	Phage λ[a]	50,000 (≤ 15)[b]	Richardson et al., 2002
	Soil	BAC (27)	3,648 (8)	Rondon et al., 2000
Biotin production	Soil / excrement enrichment	Cosmid (35)	50,000 (7)	Entcheva et al., 2001
Oxidation of polyols	Soil	Plasmid (7)	300,000 (15)	Knietsch et al., 2003b
Cation transporter	Soil	Plasmid (7)	1,480,000 (2)	Majerník et al., 2001
Cellulase	Anaerobic enrichment	Plasmid (8)	15,000 (≤ 23)	Healy et al., 1995
	Sediment enrichment	Phage λ (6)	385,000 (3)	Rees et al., 2003
Chitinase	Seawater	Phage λ (6)	825,000 (23)	Cottrell et al., 1999
Dehydratase	Soil / sediment enrichment	Plasmid (4)	560,000 (2)	Knietsch et al., 2003a
DNAse	Soil	BAC (27)	3,648 (1)	Rondon et al., 2000
Glucovanillin metabolism	Vanilla pod enrichment / soil	Plasmid (5)	205,000 (9)	***Chapter 7***
Hemolysis	Soil	BAC (45)	24,576 (29)	Rondon et al., 2000

Table 2 - continued

Activity	Analyzed habitat	Vector (average insert size, kb)	Gene bank size (positive hits)	Reference
4-Hydroxybutyrate conversion	Soil	Plasmid (7)	930,000 (5)	Henne et al., 1999
β-Lactamase	Soil	Plasmid (5)	80,000 (4)	***Chapter 2***
Lipase/esterase	Soil	Plasmid (7)	1,016,000 (4)	Henne et al., 2000
	Soil	BAC (27)	3,648 (2)	Rondon et al., 2000
	Sediment enrichment	Phage λ (6)	160,000 (3)	Rees et al., 2003
	Soil	Plasmid (10)	57,500 (≤ 117)[b]	Lorenz et al., 2001
Metalloprotease	Soil	Plasmid (10)	117,000 (1)	Lorenz et al., 2001
Oxygenase	Soil	Plasmid (7)	3,600,000 (5)	Lorenz et al., 2001
Metabolite formation	Soil	BAC (63)	12,000 (4)	MacNeil et al., 2001
	Soil	Cosmid[a]	700,000 (65)	Brady and Clardy, 2000
	Soil	Shuttle cosmid[a]	5,000 (5)	Courtois et al., 2003
	Soil	BAC (42)	3,648 (1)	Rondon et al., 2000
	Soil	BAC (44.5)	24,546 (3)	Gillespie et al., 2002

[a] Insert sizes were not reported.
[b] Since amplified libraries were screened, some of the clones may be identical. No evidence was provided that unique clones were obtained.

80,000 clones needed to be analyzed to find such an enzyme in a plasmid or a phage library, respectively (Table 2). Of course, success rates are very much determined by the microbial community under investigation and cannot be directly compared between different studies. In the recovery of lipases, for instance, the opposite situation to the amylase screening was found, with a plasmid-based cloning system being significantly more efficient than a BAC expression system.

In general, the success of a metagenome screening project depends on a number of different factors such as (a) the choice of a suitable sample for DNA extraction, (b) the availability of a DNA extraction protocol that yields sufficient quantities of high-quality DNA, (c) the use of a host-vector system that allows efficient cloning and high-level expression, and (d) a screening method of reasonable throughput. Before being able to harvest biocatalysts from the metagenome, a number of experimental hurdles thus needs to be overcome, which will be a main subject of this thesis.

Albeit to a lower extent than random cloning, PCR-based cloning methods are also being employed to recover novel enzymes. In most cases, degenerate primers are used, hybridizing with conserved regions that preferentially are located close to the extremities of the target genes. Although this approach is more conservative than the first one, new (partial) genes encoding monooxygenases (Holmes, 1999), dehalogenases (Gray et al., 2003; Marchesi and Weightman, 2003), xylanases (Radomski et al., 1988), dioxygenases (Yeates et al., 2000), and polyketide and formyltetrahydrofolate synthases (Salmassi and Leadbetter, 2003; Seow et al., 1997) were discovered in relatively small gene libraries. For instance, only 91 clones needed to be screened by restriction analysis and sequencing to discover 8 different types of formyltetrahydrofolate synthases from a termite gut microbial community (Salmassi and Leadbetter, 2003), whereas random environmental gene libraries usually need to contain 10^4 to 10^6 transformants to yield a similar number of new activities (Table 2). However, the design of suitable PCR primers is not possible for heterogeneous groups of enzymes, and PCR amplification often does not recover complete genes. Another drawback of this way of enzyme discovery is its inherent limitation to biocatalysts that share some sequence identity with already known ones.

Recently, a PCR-based sequence-independent approach was described by Stokes et al. (2001). This method makes use of PCR primers targeting a 59-base recombination site that is present in diverse bacterial taxonomic groups and that flanks gene cassettes associated with integrons. The recovered gene cassettes contained complete open reading frames, most of which did not show homology to any database entry, and which potentially encode enzymes of biotechnological interest due to the selective pressure exerted on mobile genetic elements. This study shows how methodologies for cultivation-independent enzyme discovery may evolve in the future, making use of the rapidly growing knowledge on the organization and evolution of microbial genomes.

Besides expression-based identification of biocatalysts, also large-scale shotgun sequencing projects and *in silico* identification of enzyme-coding regions are currently carried out, for instance by The Institute of Genomic Research (TIGR) and the Monterey Bay Coastal Ocean Microbial Observatory (http://www.tigr.org/tdb/MBMO/) on marine picoplancton (Béjà et al., 2000 and 2002). While this approach leads to fundamental insight into the genetic organization and function of non-

cultivated organisms and accumulates valuable sequence information in the databases, shotgun sequencing remains a rather cumbersome and expensive approach when aimed at enzyme discovery due to its indirect nature.

3. Enzyme improvement by mutagenesis techniques

Currently, there are three major approaches in enzyme engineering: (1) rational design by site-directed mutagenesis techniques; (2) random introduction of point mutations, e.g. by error-prone PCR, mutator strains, UV or chemical means; and (3) recombination of orthologous genes by gene shuffling techniques. While site-directed mutagenesis is the most conservative method, it is a powerful tool for the identification of residues involved in catalysis and substrate specificity. Often, studies based on this approach have led to improved mutants, for instance with respect to conversion rates, substrate range, or stability (for a recent review see Kazlauskas, 2000). However, the rational design of biocatalysts requires detailed knowledge about the enzyme, preferentially including a crystal structure or at least a detailed model. Even if such information is available, it is difficult to predict or rationalize the effects of a certain mutation, e.g. a change in reaction rate or enantioselectivity. This is especially the case if the mutated amino acid residues are far away from the active site of the enzyme. Once a specific residue has though been identified to play an important role in catalysis, it can be randomized to any of the 20 amino acids, using site-saturation mutagenesis. Screening of the resulting gene banks has often revealed unexpected mutants to exhibit the strongest improvements (e.g. Helfand et al., 2002; Sio et al., 2003; Taguchi et al., 2000).

In contrast to site-directed mutagenesis, the completely random introduction of point mutations has the advantage that no knowledge about the structure or mechanism of the enzyme of interest is required. Important residues, previously not known to play a role in catalysis, have been identified with this approach (e.g. Leemhuis et al., 2003; Lingen et al., 2003; Minagawa et al., 2003; Otten et al., 2002). Due to statistical reasons, however, random mutations tend to be scattered over the outer shell of the enzyme and are not focused close to the active site where they usually have a larger impact on the catalytic performance. Therefore, frequencies of mutant enzymes with significantly altered properties are expected to be relatively low in libraries generated by random mutagenesis. On the other hand, mutations in the active site are often too drastic and lead to loss of activity, in which case the bias towards distant mutations would obviously be desirable. In any case, the overall sequence diversity of enzyme variants obtained by random mutagenesis is rather limited, typically with changes in only a few amino acids between different mutants.

By random *in vitro* recombination of point mutants of the same gene or two or more different homologous genes (family shuffling), the diversity in the resulting gene bank can be enormously increased. In most cases, the properties of the resulting recombinants are difficult to rationalize as they are probably due to cumulative effects of different mutations (Crameri et al., 1998). Although gene shuffling techniques currently constitute the most powerful tool for the creation of diverse sequences and

activities (Zhao et al., 2002), their success relies on appropriate selection and screening methods that can cope with large numbers of clones. Even if these methods are available, the preparation and assaying of large gene banks remains a toilsome task.

Ideally, the creation of sequence diversity should thus be targeted to sites or regions that have a significant effect on the catalytic performance of the enzyme and ignore less important residues. In this way, the total number of possible mutants would be decreased, while the portion of potentially interesting variants is expected to rise significantly. An example of this strategy used for the improvement of a new penicillin acylase is presented in *Chapter 6.*

4. The bottleneck: screening of large expression gene banks

Environmental enzyme discovery and random improvement techniques share the necessity to screen large numbers of clones. Simple but specific enzyme assays are needed, preferentially based on substrates that cause an easily detectable signal upon conversion. The best screening methods in this respect exist for hydrolytic enzymes, which is why most enzymes recovered from environmental gene banks to date belong to this class of enzymes (Table 2).

In general, enzyme activity can either be assayed on a clone-by-clone basis, distinguishing positive and negative clones, for example, by a color reaction (screening), or by direct selection, using growth conditions that only allow the proliferation of cells exhibiting the desired enzymatic activity. Screening can be carried out in miniaturized liquid cultures, which requires the transfer of colonies to 96- or 384-well microtiter plates (MTPs) and growth in a suitable medium. Subsequently, culture aliquots are transferred to fresh plates to perform the actual enzyme assays that might involve centrifugation steps and cell permeabilization/lysis besides mixing with the chosen substrates. Plenty of instrumental assays have been developed for the high throughput screening in these formats, mainly based on spectroscopic detection of specific products by UV/VIS or fluorescence methods, IR-thermography, capillary electrophoresis or, in a somewhat lower throughput, by high-performance liquid chromatography (HPLC) and mass spectrometry (for a review see Wahler and Reymond, 2001). Recently, also a screening method for dehalogenases based on the detection of chloride release by a 96-pin halide sensor has been devised (Buta et al., 2003; http://www.igb.fhg.de/WWW/GF/Biokatalyse). Due to the work involved in colony picking and handling of hundreds to thousands of MTPs, however, expensive investments in robotic devices and spacious storage facilities (-80°C freezers) are generally required.

The use of amplified gene banks that contain multiple copies of each original clone and that can be stored as plasmid preparations in single tubes combined with screening on agar plates evades most of the above-mentioned problems. Depending on the assay used, between 500 and several thousand clones can be tested on a single square agar plate (10×10 cm in size) and positives can generally be identified by eye due to their different phenotype compared to the majority of non-active clones. Suitable assays of

this type were often adapted from classical microbiological strain characterization, including tests for some of the major groups of industrially relevant enzymes such as esterases/lipases, proteases, amylases, and cellulases (Rondon et al., 2000). In these assays, insoluble test substrates, such as tributyrin for lipases or skim milk for proteases, are added to the agar medium and active *E. coli* colonies are detected by surrounding clearing zones due to the degradation of the substrates. Also characteristics such as antimicrobial activity or the ability to cause hemolysis, which are often due to the expression of complete enzymatic pathways, can be tested by similar assays. While hemolysis is tested on blood agar, the production of antibiotics is detected by clearing zones caused by the lysis of bacterial indicator cells (e.g. *Bacillus subtilis*) that are supplied in a top agar layer.

An alternative to colony screening is the analysis of single cells using function-based cell sorting techniques. This approach is suitable for target enzymes that can be displayed on the cell-surface of their host cell, as described by Olsen et al. (2000) for mutants of the OmpT protease. Here, active clones could be detected and isolated by fluorescence-activated cell sorting (FACS), using a substrate that contains a fluorophore and a quenching group. Active cells were labeled upon enzymatic release of the quenching group because the positively charged fluorescent product remains associated to the negatively charged cell surface. In this way, hundred thousands of clones could be assayed per day.

The most convenient way of assaying gene libraries, however, is the growth selection of positive clones. With this approach, a virtually unlimited number of clones can be tested on a single agar plate, since only transformants exhibiting the desired activity will be able to form colonies. A nice example for this way of enzyme activity detection is the complementation of the leucine auxotrophy of an *E. coli* host strain by hydrolysis of phenylacetyl-L-leucine due to the action of a heterologous (penicillin) amidase (Forney and Wong, 1989; *Chapter 4*). Other selection methods, such as those described for biotin biosynthesis operons (Entcheva et al., 2001), $Na^+(Li^+)/H^+$ antiporters (Majerník et al., 2001), and chorismate mutases (Kast et al., 1996) also make use of *E. coli* strains bearing suitable auxotrophies or deficiencies. Further selectable traits are, for instance, resistance to an antibiotic or growth on an unusual substrate that constitutes the selecting carbon or nitrogen source. However, the possibility to use a selection method is always strongly determined by the metabolism of the surrogate host organism, which sometimes makes it impossible to find a suitable protocol. In some situations, this problem can be solved by adapting the characteristics of the host strain to the envisaged screening strategy, using metabolic engineering. For example, Knietsch et al. (2003b) constructed an *E. coli* strain that carries the pathway of *Citrobacter freundii* for anaerobic glycerol breakdown except for a dehydratase, allowing for the selection of clones carrying a complementing enzyme under anaerobic conditions.

A very elegant and general approach of host strain engineering has been developed in the group of de Lorenzo (de Lorenzo and Mohn, 2003; Garmendia et al., 2001). Here, the host cells are transformed with two plasmids, one carrying the environmental DNA fragment and another encoding a transcriptional activator (XylR or derivatives) and a reporter gene preceded by the corresponding inducible promoter. In order to be

active, the regulator protein needs to bind to a given (aromatic) effector molecule, which is only formed by the action of a heterologous target enzyme or pathway. The induced transcription activator then triggers the expression of the growth-conferring reporter enzyme. This enzyme could, for example, be an amidase that allows a leucine-auxotroph host cell to use phenylacetyl-L-leucine as a sole source of leucine (see above). Due to the diversity of transcription regulators, this strategy may be applicable to a wide range of reactions that do not directly lead to a selectable phenotype. The QUEST (querying for enzymes using the three-hybrid system) method of Firestine et al. (2000) works in a similar way by coupling substrate turnover to a transcriptional event. The drawback of these approaches is the need for (engineered) transcription factors that specifically interact with the substrate or the reaction product and do not cross-react with other compounds of the reaction mixture. Similar problems arise when monitoring enzymatic reactions with immunoassays, since specific antibodies are needed, which requires lengthy immunization experiments. If those antibodies are available, however, sophisticated and highly efficient screens can be carried out, such as by Taran et al. (2002) for improved enantioselectivity in the conversion of benzoyl formic acid to mandelic acid.

5. Penicillin acylase – an enzyme of industrial interest

Driven by the need to develop improved processes for the production of modified penicillin and cephalosporin antibiotics, Cole suggested the use of penicillin acylase of *E. coli* to catalyze their synthesis (Cole, 1969). The production of semi-synthetic β-lactam antibiotics had become necessary since bacterial resistance against penicillin G emerged rather quickly after its introduction in the 1940's. Penicillin G selectively interferes with bacterial cell wall synthesis by acting as a suicide substrate for transpeptidases (penicillin-binding proteins) that are involved in the cross-linking of the peptidoglycan layer surrounding the inner cell membrane of bacteria. To prevent the action of penicillin G, many bacteria have acquired β-lactamases that hydrolyze the amide bond in the β-lactam ring and release inactive penicillinoic acid. This is the major mechanism by which organisms become resistant to penicillin G and other β-lactam antibiotics (Essack, 2001) (Fig. 2). Another way to obtain resistance is the production of penicillin-binding proteins that have a low affinity for β-lactam antibiotics (Spratt, 1994).

Due to the rapid spread of β-lactamase genes between different bacterial species and because of the need to improve the treatment of emerging bacterial pathogens, new antibiotics constantly need to be developed. Fortunately, small changes in the structure of the penicillin G molecule are often sufficient to escape modification by existing β-lactamases, while preserving its anti-infective activity or even broadening the antibacterial spectrum. For instance, the introduction of an amino group on the Cα-position led to the first semi-synthetic penicillin that is active against Gram-negative bacteria: ampicillin. Nowadays, semi-synthetic antibiotics such as ampicillin,

amoxicillin, cephalosporin or cefadroxil belong to the world's major pharmaceutical products with a market of about 15 billion € (Elander, 2003).

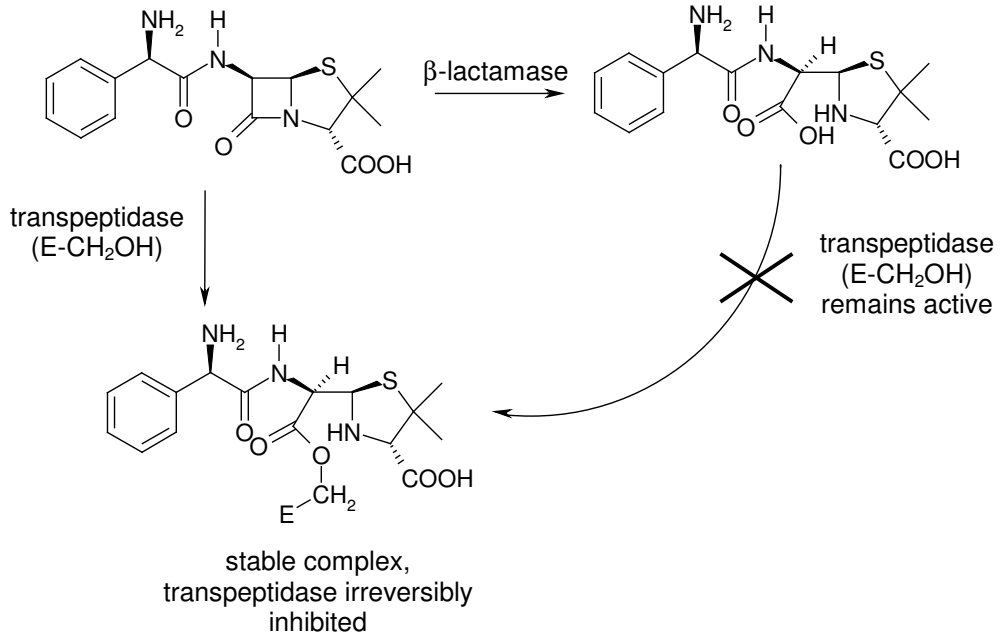


Figure 2. Mode of action of penicillin G and bacterial resistance due to β-lactamase activity. Penicillin G inactivates the transpeptidase, which is involved in cross-linking peptidoglycan chains, by causing the formation of a stable transpeptidase-penicilloyl complex. As a consequence, a stable cell wall with an intact peptidoglycan layer cannot be established and cells lyse during proliferation. When a β-lactamase is present, however, the β-lactam moiety can be hydrolyzed to penicillinoic acid that cannot inhibit the transpeptidase.

The starting material for the production of semi-synthetic antibiotics is mostly penicillin G, which is obtained by fermentation using the fungus *Penicillium chrysogenum*. The annual production exceeds 40,000 tons. About 75 % of this amount is used for the production of 6-aminopenicillanic acid (6-APA) (Elander, 2003), the β-lactam nucleus that is used to synthesize semi-synthetic penicillins (Fig. 3). When production of semi-synthetic β-lactam antibiotics started in the late 1950's, the whole process from splitting of penicillin G to coupling of the β-lactam nucleus to a modified side-chain was carried out by purely chemical methods. This route required the use of hazardous chemicals, organic solvents, and a high energy input, as well as plenty protection and deprotection steps, resulting in high costs and environmental problems. As a consequence, interest in enzymatic routes that can be run at moderate temperatures in aqueous media increased. Already in 1960, a group of enzymes called penicillin acylases (EC 3.5.1.11) was discovered that can hydrolyze penicillin G and release 6-APA under mild conditions (Rolinson et al., 1960). Nowadays, the large majority of 6-APA is produced enzymatically (Bruggink et al., 1998).

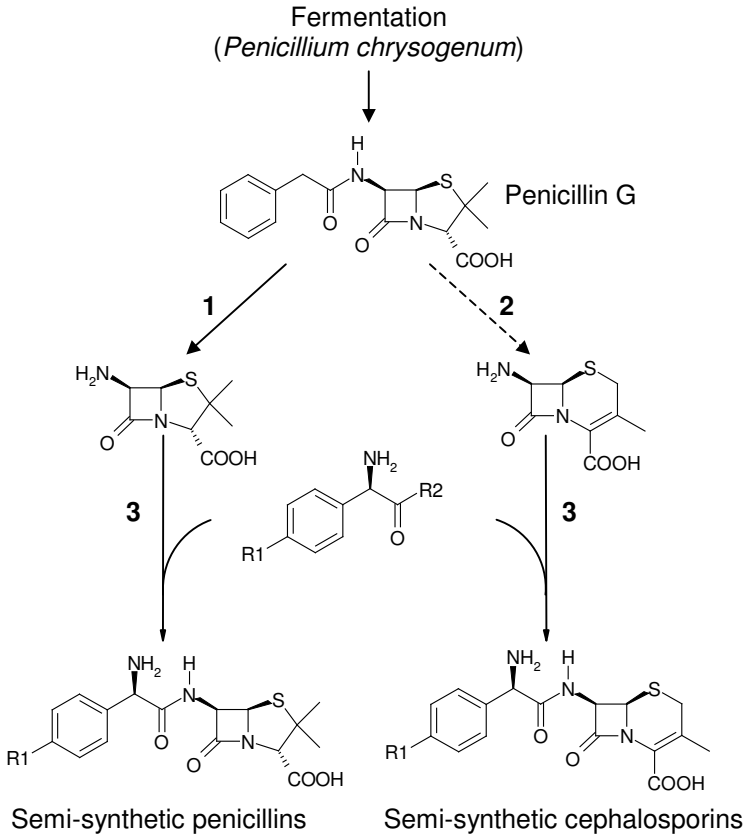


Figure 3. Production of semi-synthetic β-lactam antibiotics from penicillin G. (1) Hydrolytic release of 6-APA. (2) Multi-step chemical ring-expansion and hydrolytic release of 7-ADCA. (3) Condensation of a β-lactam moiety and a modified side-chain molecule. In the kinetically controlled enzymatic reaction, activated acyl-donors are used, mostly amide (R2 = NH_2) or methyl ester (R2 = OCH_3) derivatives of e.g. D-phenylglycine (R1 = H) or *p*-hydroxy-D-phenylglycine (R1 = OH). This leads to the formation of semi-synthetic penicillins such as ampicillin (R1 = H) or amoxicillin (R1 = OH) and semi-synthetic cephalosporins such as cephalexin (R1 = H) or cefadroxil (R1 = OH).

Penicillin acylases cannot only hydrolyze β-lactam antibiotics, but also synthesize them when reactive side chain precursors (acyl donors) are employed (Fig. 4). In this case, the synthesis of the antibiotic is kinetically controlled because of two side reactions that can occur: (1) the hydrolysis of the activated acyl donor, and (2) the hydrolysis of the formed antibiotic. As a consequence, the use of penicillin acylase in an antibiotic synthesis reaction leads to a temporary accumulation of the desired product that, after reaching a maximum, decreases to an unfavorable equilibrium concentration. Obviously, enzymes showing a high preference for the antibiotic synthesis reaction compared to the hydrolytic side reactions are the most suited to catalyze this kind of conversion. Several processes using *E. coli* ATCC11105 penicillin acylase for the synthesis of semi-synthetic antibiotics have been described (Bruggink et al., 1998). Unfortunately, due to its kinetic characteristics, the wild-type *E. coli* enzyme is far from being ideal. However, its synthetic performance could be improved by structure-inspired site-directed mutagenesis. Based on the crystal structure solved by Duggleby et al. (1995) and a structure of the βN241A mutant complexed with

penicillin G (Alkema et al., 2000), several residues besides the catalytic amino acids were identified that could have a strong impact on the synthetic performance of the enzyme. Especially residues αR145 and αF146 located on the 23 kD α-subunit and βF24 of the 63 kD β-subunit of the heterodimeric enzyme play important roles in substrate binding and in modulating the hydrolytic and acyl-transfer activity of the enzyme (Alkema et al., 2002b and 2002c). By replacing the native residues, single mutants were created that, under laboratory conditions, allow up to 2.5 times higher yields in semi-synthetic β-lactam antibiotic synthesis than the wild-type enzyme. Unfortunately, the activity of the respective mutants was strongly decreased, typically to less than 10 % of the wild-type activity.

The provision of biocatalysts that combine excellent synthetic properties with high activity thus remains a challenging task and a prerequisite for the development of processes that are competitive with traditional chemical condensation methods.

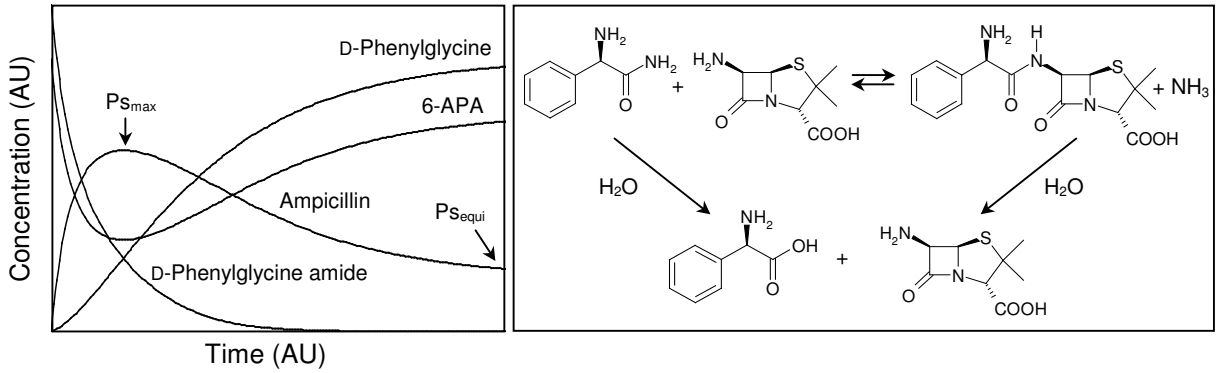


Figure 4. Kinetically controlled synthesis of semi-synthetic antibiotics. Left: Typical progress curve of antibiotic (here: ampicillin) synthesis. The maximal yield of antibiotic Ps_{max} is transiently higher than its equilibrium concentration. Right: The two hydrolytic side-reactions that occur during antibiotic synthesis (shown for ampicillin production).

6. Scope of this thesis

In the present thesis, different aspects of metagenome mining and biocatalyst development are explored. The challenge of obtaining penicillin acylases with improved synthetic performance and altered substrate specificities caused us to focus on the isolation and improvement of new members of this class of enzymes from the environmental gene pool. Besides, also more general aspects, such as the accessibility of the metagenome by random cloning as well as efficient ways of library construction and high-throughput screening/selection, were investigated.

Chapter 2 is concerned with the assessment of different methods for environmental DNA extraction. A comprehensive comparison of protocols based on DNA recovery directly from environmental samples (direct DNA extraction) and DNA isolation from cells that are firstly recovered from the sample matrix (indirect DNA

extraction) is presented. Our analysis showed that DNA extracted by indirect methods is more suited for library construction due to its better quality, larger bacterial diversity, and lower eukaryotic DNA content, although smaller amounts of DNA are generally recovered than with direct DNA extraction protocols.

In **Chapter 3**, the question to which extent the metagenome can be exploited by random expression cloning is tackled. Functional screening requires the expression of active enzymes in a heterologous host. Therefore, only a fraction of all environmental genes can be expected to be readily recovered. Based on statistical considerations and using a bioinformatics approach, we quantified the part of the metagenome that is likely to be expressed in *E. coli*, and we determined the optimal insert size for expression cloning of single genes or small operons.

Chapter 4 describes the construction of metagenomic gene banks from different DNA sources, such as loam soil and enrichment cultures, and the selection of amidase-expressing clones. Various active clones were recovered, displaying distinct substrate profiles and carrying diverse insert sequences. Most of these clones carried open reading frames that were not similar to any known enzyme, indicating that we may have isolated completely new types of proteins with amidase activity.

Chapter 5 is dedicated to a detailed kinetic characterization of a new penicillin acylase, termed PAS2, that was recovered from one of the gene libraries described in *Chapter 4*. The new enzyme allowed clearly improved yields in the synthesis of 6-APA derived semi-synthetic antibiotics.

The further improvement of PAS2 by semi-random function-inspired mutagenesis is presented in **Chapter 6**. Here, three conserved active site residues, known to be involved in substrate binding in the penicillin acylase of *E. coli*, were simultaneously randomized by site-saturation mutagenesis. This approach led to the isolation of a number of significantly improved biocatalysts. The kinetics of the three best mutants were determined to rationalize the observed increases in product yields.

In **Chapter 7**, we focus on the isolation of representatives of another group of industrially relevant enzymes, the β-glucovanillin glucosidases. This group of enzymes is involved in the release of the natural flavor vanillin from its precursor glucoside, leading to a high value product. Although a number of putative target enzymes could be recovered, the chosen selection method also led to the recovery of false-positive clones. We use this example to discuss some general pitfalls of growth selection methods.

In the last chapter, **Chapter 8**, the obtained results are shortly summarized and discussed with respect to their importance for metagenome cloning and biocatalytic application. Furthermore, some recent developments that may facilitate the harvest of novel biocatalysts from the metagenome are briefly discussed.

2

Efficient recovery of environmental DNA for expression cloning by indirect extraction methods

Esther M. Gabor, Erik J. de Vries, and Dick B. Janssen

FEMS Microbial Ecology, 2003, **44:** 153-163

Using direct and cell extraction-based (indirect) isolation methods, DNA was obtained from environmental samples with largely differing characteristics (loam soil, sand soil, sediment, activated sludge, and compost) and evaluated with respect to the comprised bacterial diversity and its suitability for expression cloning in *Escherichia coli*. Indirect DNA extraction methods yielded 10 to 100-fold lower amounts of DNA than direct procedures, but the bacterial diversity of DNA recovered by indirect means was distinctly higher as shown by denaturing gradient gel electrophoresis (DGGE). Furthermore, much lower amounts of eukaryotic DNA were co-extracted if cell extraction-based methods were used (< 8 % of eukaryotic DNA by indirect methods versus 61-93 % by direct lysis protocols). Considering the higher purity, i.e. higher cloning efficiency of DNA isolated by indirect methods, similar numbers of clones carrying prokaryotic inserts could be produced by either strategy. Gene banks prepared from directly extracted DNA, however, are expected to contain large portions of clones with eukaryotic inserts, whereas those constructed from indirectly isolated DNA should mainly contain inserts of bacterial origin. As eukaryotic genetic information is generally not expressed in bacterial host organisms but increases the library size, our findings suggest that the use of indirect DNA isolation methods allows the construction of environmental gene banks of superior quality.

1. INTRODUCTION

Recovery, cloning and expression screening of environmental DNA without preceding cultivation is a recent approach to exploit the biocatalytic potential of microbial communities present in environmental samples (Rondon et al., 1999). It has been of growing interest to both microbial ecologists and to biochemists looking for novel biocatalysts as the genetic information of theoretically all indigenous bacteria can be accessed, including the predominant fraction of microorganisms that is recalcitrant to cultivation. Using this methodology, a number of previously unknown genes, in some cases encoding for entire pathways, has been isolated during the last few years (Cottrell et al., 1999; Henne et al., 1999 and 2000; Majerník et al., 2001; McNeil et al., 2001; Rondon et al., 2000; Wang et al., 2000). Although circumventing the time-consuming microbiological work implicit to cultivation-based screening techniques, the construction of environmental gene banks is still relatively laborious due to size and complexity of most microbial communities. In view of the large number of 1,000-10,000 prokaryotic species that might be present in one gram of soil (Torsvik et al., 1989), large-scale cloning techniques need to be used to cover the collective genomes, which require substantial amounts of high-quality environmental DNA.

Various protocols have been described for DNA recovery from soil and sediment samples, which can be classified as direct and indirect DNA extraction procedures. Direct DNA isolation is based on cell lysis within the sample matrix and subsequent separation of DNA from the matrix and cell debris (pioneered by Ogram et al., 1987), whereas the indirect approach involves the extraction of cells from the environmental

material prior to the lytic release of DNA (pioneered by Holben et al., 1988). Cell extraction methods are generally acknowledged to yield DNA of higher molecular weight and greater purity than direct lysis procedures. However, in many cases the amounts of DNA recovered by this strategy are significantly lower, which is why DNA for gene bank construction is commonly isolated from soils by direct lysis protocols that readily yield the required microgram amounts of DNA (Henne et al., 1999 and 2000; Majerník et al., 2001; McNeil et al., 2001; Rondon et al., 2000; Wang et al., 2000). Due to the generally higher DNA yields, direct procedures have also been assumed to access larger fractions of indigenous microbial populations and to recover nucleic acids of larger genetic diversity than indirect methods (Leff et al., 1995; Roose-Amsaleg, 2001; Steffan et al., 1988). However, a recent report dealing with the comparison of different direct lysis protocols showed that greater DNA yield does not always equal greater bacterial species richness and that sequence representation is strongly influenced by the extraction method used (Stach et al., 2001). Consequently, DNA extracts obtained by indirect methods must not necessarily be less representative for the targeted microbial community than those obtained by direct lysis.

Besides the apparent selectivity of DNA recovery methods for specific groups of bacteria, the amount of co-extracted eukaryotic nucleic acids may also bias the yield - diversity correlation. For construction of expression gene banks in *Escherichia coli* or other bacterial hosts, the eukarya content is a critical parameter of DNA extracts used. As expression in these systems is in most cases limited to prokaryotic genes, eukaryotic DNA will evade analysis and increase the number of clones that need to be prepared and screened. This negative effect is still increased by the generally much larger genome size of eukaryotes (3 to 140,000 Mb) compared to prokaryotic organisms (0.6 to 9.5 Mb) (Vellai and Vida, 1999). For example, when recovering the complete DNA from an environmental sample comprising only 0.1 % of eukaryotic cells, the resulting extract would consist of 91 % of eukaryotic nucleic acids, assuming an in average 10,000 times larger genome size of the present eukaryotes. Consequently, an environmental gene bank prepared from such DNA would need to be more than 10-fold larger to cover the same range of prokaryotic genomes than if prepared from eukarya-free nucleic acids. Although previous studies suggested that eukaryotic DNA may be present in DNA extracts prepared by direct lysis (Lloyd-Jones and Hunter, 2001; Steffan et al., 1988; Tien et al., 1999), the extent of their recovery by different direct and indirect lysis methods has not been thoroughly investigated.

In this study, we systematically compared the two DNA isolation strategies with respect to their suitability for environmental gene bank construction. Two typical protocols of each category were used for the isolation of DNA from different soils, sediment, compost material, and activated sludge. We evaluated the treatments with respect to yield, molecular weight, and cloning efficiency of the environmental DNA. A major objective was the quantification of co-extracted eukaryotic nucleic acids and the determination of bacterial diversity in the obtained DNA extracts by denaturing gradient gel electrophoresis (DGGE).

2. MATERIALS AND METHODS

2.1 Environmental samples

Single samples of about 500 g were collected of each environmental material, transferred to the laboratory and stored at –20°C within 2 hours after sampling. Marine sludge was collected from the surface of daily-inundated saline mud flats during low tide near Paesens-Moddergat, The Netherlands. Aerobic activated sludge was obtained from a municipal sewage plant (Garmerwolde, The Netherlands) and thermophilic stage compost from an urban composting facility (Groningen, The Netherlands). Sand and loam soils were sampled from the upper 5 to 15 cm of soil at the lakeshore of the Lauwersmeer (The Netherlands) and a nearby-located agricultural field, respectively. Coarse plant material and stones were removed prior to storage of the samples. Sample pH was measured in a slurry of 9 parts distilled water and 1 part sample material. Moisture contents were determined by drying 10-g aliquots at 100°C for 2 days. To determine the organic matter content, 5 to 10 g of dried sample material was incubated at 500°C for 3 h and the decrease in weight was measured. The amounts of prokaryotic cells present in the different samples were determined with a direct microscopic count procedure based on DTAF [5-(4,6-dichlorotriazin-2-yl)amino fluorescein] staining of prokaryotic cells (Bloem et al., 1994). Specimens were examined using an Olympus CK 40 epifluorescence microscope.

2.2. Direct DNA extraction protocols

2.2.1 Soft lysis method

To disrupt microorganisms by solely enzymatic and chemical means a modified protocol of Zhou et al. (1996) was used. Triplicate 1-g environmental samples were homogenized by vortexing in 750 µl lysis buffer [100 mM Tris-HCl, 100 mM sodium EDTA, 1.5 M NaCl, 1 % hexadecylmethylammonium bromide (CTAB); pH 8] at maximum speed for 5 min in 2-ml screw-cap tubes. 40 µl lysozyme (50 mg ml^{-1}) and 10 µl proteinase K (10 mg ml^{-1}) were added before incubation at 37°C for 30 min. After addition of 200 µl SDS (20 %), mixtures were incubated at 65°C for 2 h with vigorous shaking by hand every 30 min. The supernatants were collected by centrifugation at 6,000 g for 10 min at room temperature, and the pellets were re-extracted twice by adding 500 µl of lysis buffer, vortexing for a few seconds, and incubation at 65°C for 10 min. Centrifugation was carried out as before. The combined supernatants were extracted with an equal volume of chloroform before precipitating the DNA from the recovered water phase by addition of 0.6 volumes of isopropanol and overnight incubation at 4°C. The precipitates were collected by centrifugation at 16,000 g, washed with 70 % ethanol, and suspended in a total volume of 50 µl TE buffer (10 mM Tris-HCl, 1 mM sodium EDTA; pH 8).

Lysis efficiencies (LE) were calculated by {1–[cells g^{-1} (dry wt) extracted pellet]/[cells g^{-1} (dry wt) untreated sample]}×100. Moisture contents and cell numbers of the extracted pellets were determined as described for the untreated sample materials.

2.2.2. Harsh lysis method

The effects of mechanical cell disruption were studied by including a bead-milling step in the above protocol. Triplicate 1-g samples were homogenized in lysis buffer and supplied with lysozyme and proteinase K as described above. Additionally, 0.7 g zirkonia/silica beads (0.1 mm diameter, BioSpec Products) were added. After incubation at 37°C for 30 min, the mixtures were agitated at top speed on a vortex mixer for 3 min, which corresponds to the vigor of a 1-min treatment in a mini bead beater (BioSpec Products) as judged by the size distributions of recovered DNA fragments (not shown). Subsequent steps were carried out as described for the soft lysis method.

2.3. Indirect DNA extraction protocols

2.3.1 Blending method

A modified repeated blending protocol (Holben et al., 1988) was used to mechanically release bacterial cells from the sample matrix. Duplicate 50-g environmental samples were dispersed in 100 ml blending buffer (100 mM Tris-HCl, 100 mM sodium EDTA, 0.1 % SDS, 1 % CTAB; pH 8) and homogenized in a standard blender (Moulinex) for three 1-min intervals, with 1-min breaks in-between to allow cooling. Coarse particles were collected by low-speed centrifugation (1,000 g for 10 min at 10°C), resuspended in 100 ml blending buffer, and subjected to another two blending-centrifugation cycles as described above. Supernatants obtained during the three rounds of cell extraction were pooled. Direct microscopic cell counts of the combined supernatants were compared to those of the untreated sample materials to determine cell extraction efficiencies. Supernatants were centrifuged at high speed (10,000 g) for 30 min at 4°C to collect the microbial cell fraction, which was subsequently washed in 150 ml of 0.1 % sodium pyrophosphate (4°C). After a second wash in 100 ml Chrombach buffer (0.33 M Tris-HCl, 1 mM EDTA; pH 8), pellets were resuspended in 8 ml lysis buffer, 160 µl lysozyme (50 mg ml^{-1}), and 40 µl proteinase K (10 mg ml^{-1}) solution and incubated at 37°C for 30 min. Lysis was completed chemically by adding 1 ml of 20 % SDS and incubation for 2 h at 65°C with rotary shaking (225 rpm). Chloroform extraction and isopropanol DNA precipitation were carried out as described for the direct lysis methods. DNA pellets were dissolved in 250 µl TE buffer.

2.3.2 Cation-exchange method

To disperse cells by chemical means, an adapted protocol from Jacobsen and Rasmussen (1992) was used. Duplicate 50-g environmental samples were mixed with 10 g Chelex 100 (BioRad) and 100 ml extraction buffer (0.1 % sodium deoxycholate, 2.5 % polyethylene glycol 6,000) and shaken at room temperature for 1 h at 100 rpm on an orbital shaker. Mixtures were centrifuged at low speed (1,000 g) for 15 min at 10°C to remove coarse particles including the cation-exchange resin. Supernatants were transferred to fresh centrifuge tubes and subjected to high-speed centrifugation (10,000 g) for 30 min at 4°C to harvest microbial cells. Subsequent steps were

performed as described for the blending method. DNA pellets were suspended in 100 µl TE.

2.4. Total DNA yield and molecular mass

Crude DNA extracts were analyzed on 0.4 % (w/v) agarose gels that were post-run stained with ethidium bromide. Gel photographs were scanned and analyzed with the NIH image software (available at http://rsb.info.nih.gov/nih-image). Genomic DNA was quantified using a calibration curve that was prepared for each image with the five smallest fragments of the Smart Ladder molecular weight marker (Eurogentec) being present in known amounts (20 to 100 ng per band).

2.5. DNA purification

For molecular analysis and cloning experiments, replicate crude DNA extracts were pooled and purified by preparative gel electrophoresis on 3 % (w/v) agarose gels. Genomic DNA was extracted from gel with the QIAEX II gel extraction kit (Qiagen). Care was taken to extract all DNA bands visible on the gel to ensure also the recovery of plasmid borne DNA. Recovery was routinely > 80 %.

Table 1. Small subunit rRNA gene probes and PCR primers used in this study

Probe/ primer	Specificity	Sequence (5'→3')	Target site[a]	T_h[b] (°C)	**Reference**
UNIV1390	Universal	GAC GGG CGG TGT GTA CAA	1390-1407	45	Zheng et al., 1996
EUB341	Bacteria	CCT ACG GGA GGC AGC AG	341-357	45	Muyzer et al., 1993
ARC915	Archaea	GTG CTC CCC CGC CAA TTC CT	915-934	53	Stahl and Amann, 1991
EUC502	Eukarya	ACC AGA CTT GCC CTC C	502-517	39	Amann et al., 1990
U968	Bacteria	AAC GCG AAG AAC CTT AC	968-984	n. a.[c]	Engelen et al., 1995
L1401	Bacteria	CGG TGT GTA CAA GAC CC	1385-1401	n. a.[c]	Engelen et al., 1995

[a] *E. coli* numbering
[b] T_h, probe-specific hybridization temperature
[c] n. a., not applicable

2.6. Quantitative dot blot hybridization

Four digoxigenin-labeled oligonucleotide probes (Eurosequence, The Netherlands) were used to trace the origin of environmental nucleic acids (Table 1). Purified DNA was denatured at 95°C for 10 min, and 1-1.5 µg aliquots were spotted in triplicate on positively charged nylon membranes (Roche). Nucleic acids were cross-linked to the air-dried membranes by baking at 120°C for 30 min. Membranes were prehybridized for 2 h at the probe-specific hybridization temperature T_h in a buffer consisting of 5×

SSC (SSC is 150 mM NaCl and 15 mM sodium citrate; pH 7), 0.02 % (w/v) SDS, 0.1 % (w/v) N-lauroyl sarcosine, and 1 % Blocking Reagent (Roche). Hybridization was carried out overnight at T_h in 15 ml of the same buffer supplemented with 200 pmol of the probe. Membranes were washed twice with 2× SSC, 0.1 % SDS for 5 min at room temperature and twice with 0.5× SSC, 0.1 % SDS for 15 min at T_h. Chemoluminescent detection of the hybridized probe by anti-DIG-alkaline phosphatase Fab fragments and CSPD (Disodium 3-(4-methoxyspiro {1,2-dioxetane-3,2'-(5'-chloro) tricyclo [3.3.1.1^{3,7}] decan}-4-yl)phenyl phosphate] and subsequent removal of bound probe was carried out according to the instructions of the manufacturer (Roche). Membranes were successively hybridized to all four probes as described above. X-ray films were digitally photographed and analyzed using the NIH image software. Bacterial, archaeal, and eukaryotic response factors of the four probes were determined with twofold dilution series of genomic DNA (6.4 µg to 100 ng) of *Bacillus megaterium*, *Pyrococcus furiosus*, and *Hansenula polymorpha*.

2.7. PCR

Primers U968 and L1401 (Table 1) were used to amplify a 402-bp section of bacterial 16S rRNA genes, including the highly variable V6 region (Engelen et al., 1995). A GC clamp (5'-CGC CCG GGG CGC GCC CCG GGC GGG GCG GGG GCA CGG GGG G-3') was attached to the 5' end of primer U968 to confer melting stability to the PCR products during denaturing gradient gel electrophoresis (DGGE). Specific amplification of the target sequences was routinely achieved by using 1-10 ng of template DNA in a total volume of 80 µl PCR reaction mixture [300 µg ml^{-1} BSA, 1.25 nmol ml^{-1} of each primer, 200 µM of each dNTP, 1× PCR buffer, and 25 U ml^{-1} of *Taq* polymerase (Roche)]. After an initial denaturation step of 3 min at 94°C, PCR temperature cycles of 1 min of denaturation at 94°C, 1 min of annealing, and 1 min of primer extension at 72°C were performed. During 10 initial touchdown cycles, the annealing temperature was lowered from 56 to 47°C in steps of 1°C per cycle. Subsequently, 25 cycles were done at 46°C followed by a final extension step of 10 min at 72°C. PCR products were purified on 0.8 % (w/v) agarose gels using the QIAquick gel extraction kit (Qiagen).

2.8. DGGE analysis

DGGE was essentially performed as described previously (Muyzer et al., 1993). A device manufactured by Ingeny International BV (The Netherlands) was used to separate PCR products on 9 % polyacrylamide gels containing a 30-70 % gradient of urea and formamide [100 % solution: 40 % (v/v) formamide, 7 M urea]. Samples containing 100-150 ng of PCR products were loaded per gel slot and run for 16 h at a constant voltage of 110 V and a temperature of 60°C. Gels were silver stained (Bassam et al., 1991) and digitally photographed. Pixel density profiles were obtained by using the ImageQuaNT software version 4.1b (Molecular Dynamics) that allowed the determination of exact position, number (richness *S*) and intensity (relative abundance p_j) of individual bands. Shannon-Weaver diversity indices were calculated by $H' = -\sum p_j \log_2 p_j$ (Pielou, 1977).

2.9. Sequencing of DGGE bands

Small amounts of acrylamide containing the DNA fragments of interest were scraped from gel by use of a sterile needle which was subsequently immersed in 50 µl of PCR reaction mixture without BSA. PCR products were prepared as described above and cloned into pCR4-TOPO according to the instructions of the manufacturer (Invitrogen). Inserts of at least three clones per DGGE band were sequenced at the Medical Biology Department of the University of Groningen to test for unspecific PCR amplification products (background) and whether different 16S rRNA gene segments were present. Sequences were compared to the 16S rDNA database of the Ribosomal Database Project RDP-II (http://rdp.cme.msu.edu/html) by using the Sequence Match service (Maidak et al., 1999), and to GenBank entries by using BLAST software (Altschul et al., 1990).

2.10. Cloning

The purity of DNA extracts was estimated in terms of blunt-end cloning efficiencies (transformants per µg of environmental DNA fragments) in the *Eco*RV site of the high-copy plasmid vector pZero-2 (Invitrogen). Environmental DNA fragments of 4 to 6 kb were prepared by mechanical shearing of DNA using a nebulizer (Invitrogen), and subsequent blunting with Klenow and T4 polymerase according to the instructions of the manufacturer (Roche). Ligation using a 1:10 vector:insert ratio and transformation to *E. coli* TOP10 cells [Δ(ara-leu)7697] by electroporation was done as described in the manual provided with pZero-2, using a Gene Pulser apparatus (BioRad). Transformants were spread on Luria-Bertani (LB) agar medium [15 g l^{-1} agar (Difco)], containing 10 mg l^{-1} kanamycin for the selection of transformants carrying a pZero-2-derived recombinant plasmid. After incubation at 37°C for 1 day, colony forming units (cfu) were enumerated. At least 20 transformants were separately grown overnight in 5 ml LB medium supplied with 10 mg l^{-1} kanamycin. Plasmid DNA was isolated from these cultures using the High Pure Plasmid Isolation kit (Roche), and checked for insert size by running *Bam*HI/*Xho*I (Roche) digested samples on a 0.8 % agarose gel and comparing bands to the Smart Ladder molecular weight standard (Eurogentec). Average insert sizes were found to be 5.2 kb in all gene banks, with background levels of self-ligated vector molecules of ≤ 5 %.

2.11 Gene bank screening

Transformants obtained from loam soil DNA (blending method) were washed from LB agar plates with minimal medium [per liter: 5.3 g $Na_2HPO_4 \cdot 12H_2O$, 1.4 g KH_2PO_4, 0.2 g $MgSO_4 \cdot 7H_2O$, 1.0 g $(NH_4)_2SO_4$, 1 ml vitamin solution (Janssen et al., 1984), and 5 ml of a trace element solution; the trace element solution contained per liter 780 mg $Ca(NO_3)_2 \cdot 4H_2O$, 200 mg $FeSO_4 \cdot 7H_2O$, 10 mg $ZnSO_4 \cdot 7H_2O$, 10 mg H_3BO_3, 10 mg of $CoCl_2 \cdot 6H_2O$, 10 mg $CuSO_4 \cdot 5H_2O$, 4 mg $MnSO_4 \cdot H_2O$, 3 mg $Na_2MoO_4 \cdot 2H_2O$, 2 mg $NiCl_2 \cdot 6H_2O$, and 2 mg $Na_2WO_4 \cdot 2H_2O$]. Aliquots from this amplified gene bank were spread on a number of different agar media containing either selective or indicator agents, and incubated at 30°C for 3 days before scoring the positive clones. All media contained kanamycin at a concentration of 10 mg l^{-1}. Amidase-expressing clones were

selected on minimal medium plates [15 g l^{-1} MP agarose (Roche)] supplied with 0.2 % (w/v) glucose and 10 mg l^{-1} phenylacetyl-L-leucine as a sole source of leucine (Forney and Wong, 1989). Selection for β-glucosidase encoding transformants was carried out on minimal agar medium (15 g l^{-1} agar) plus 5 mg ml^{-1} leucine and 2 mM β-D-glucovanillin as the only source of carbon. Clones with β-lactamase activity were selected on LB agar plates, containing 50 mg l^{-1} of ampicillin. The presence of amylase activity was indicated by a bright orange halo when flooding replicate Bacto Starch agar plates with Bacto Stabilized Gram Iodine (Difco) after 3 days of growth. Lipase activity was tested on Spirit Blue agar supplied with Lipase Reagent (Difco) with a clear halo indicating positive transformants. Due to the use of an amplified gene bank, active clones obtained in each screen needed to be analyzed by enzymatic restriction to determine the number of unique transformants exhibiting the respective enzyme activity.

3. RESULTS AND DISCUSSION

3.1. Isolation of DNA from environmental samples

In the last fifteen years, a vast number of protocols for DNA extraction from environmental samples has been published and even commercial soil DNA extraction kits have become available (e.g. from Bio101 La Jolla, CA, USA; Borneman et al., 1996). These kits as well as most of the other published methods have improved the original direct DNA extraction procedure of Ogram et al. (1987) mainly in terms of DNA yield and ease of use by varying the lysis conditions in many ways. However, the basic concept of cell lysis by enzymatic (lysozyme) and/or hot detergent (SDS) treatment is still the core of many recent DNA extraction methods (Kresk and Wellington, 1999; Rondon et al., 2000; Zhou et al., 1996). Besides, various protocols make use of mechanical forces created by e.g. bead beating, freeze-thawing or grinding to disrupt more rigid cell structures. We therefore chose a classical soft lysis protocol as well as a harsh lysis procedure including the widely used bead beating (Borneman et al., 1996; Kresk and Wellington, 1999; Kuske et al., 1998; Miller et al., 1999) to represent the direct approach of DNA extraction (Fig. 1).

Compared to direct methods, indirect DNA extraction methods are less diverse, all relying on the release of microorganisms from the sample matrix by either mechanical or chemical means (or a combination of both; Hopkins et al., 1991) and their subsequent collection by differential or density gradient centrifugation. Blending or rotating pestle homogenization of loam soil have been described to be most efficient in mechanical soil dispersion (Lindahl and Bakken, 1995). While the addition of Chelex 100 did not improve cell yields in these cases, cation-exchange resins have proven to be useful in soil dispersion on their own (Edwards and Bremner, 1965).

To reveal possible differences between the two dispersion strategies, a standard blending protocol and a method solely based on the use of Chelex 100 were selected (Fig. 1). In both cases, bacterial cells were recovered by differential centrifugation, which allows higher DNA yields than the collection of cells in a density (e.g.

Nycodenz) gradient although the purity of DNA extracts may be lower (Lindahl and Bakken, 1995).

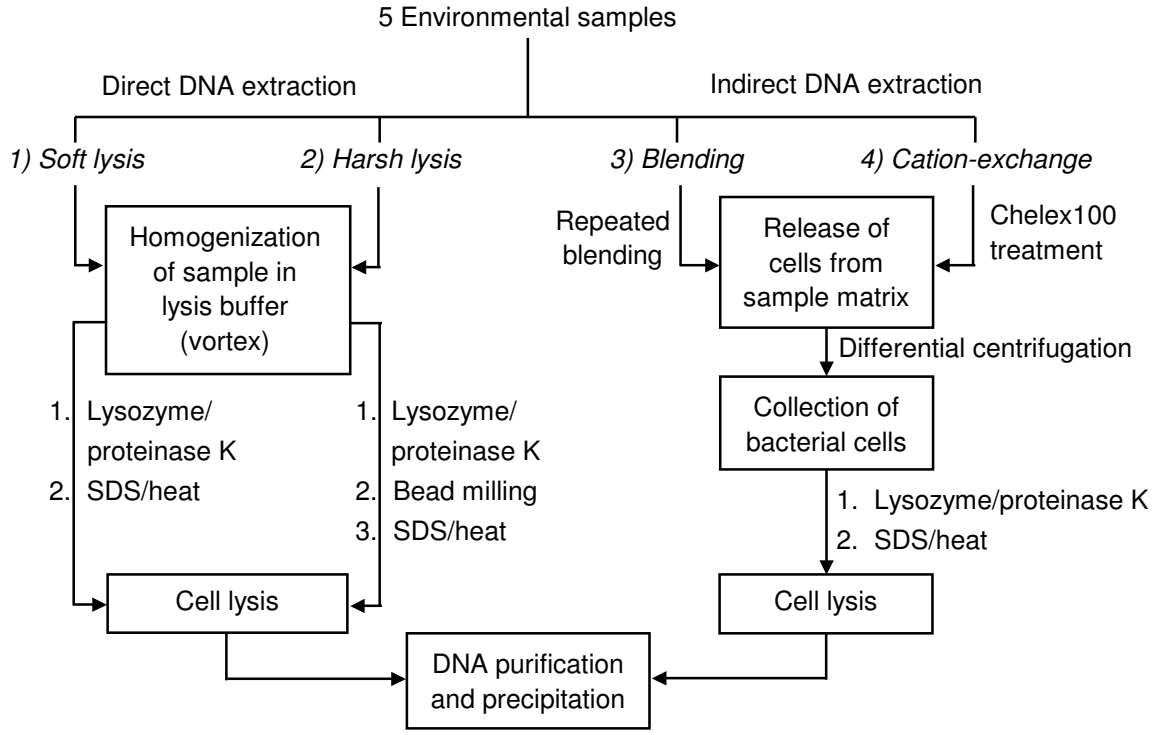


Figure 1. The four DNA recovery protocols used in this study. To emphasize the effects of the different methods, all protocols comprised the same basic procedure for cell lysis (lysozyme, proteinase K, and hot SDS treatment) followed by chloroform extraction and isopropanol precipitation of the released DNA.

All four protocols were used to recover DNA from five different environmental samples (Table 2). The amounts of DNA isolated by direct lysis were 10- to 100-fold higher than for protocols based on cell extraction (Table 3). Activated sludge and loamy sand allowed highest recoveries with all protocols, corresponding to their high indigenous cell numbers. Although purely enzymatic and chemical lysis methods are regarded as not sufficiently vigorous to lyse a large number of microorganisms (Kuske et al., 1998; Miller et al., 1999), soft direct lysis outperformed the harsh treatment by recovering up to 2 times more DNA from most environmental samples. Whereas Jacobsen and Rasmussen (1992) found about equal DNA yields from a seeded soil by both indirect methods, our results show 2 to 8-fold higher DNA recoveries with blending. This is in agreement with the higher cell extraction efficiencies reached by this treatment for all environmental samples studied.

Table 2. Characteristics of the environmental samples used in this study

Sample	pH	Moisture content [%]	Organic matter content [%][a]	Cells g^{-1} (wet wt) of sample[b]
Marine sludge	8.0	52	10.8	$(4.79 \pm 0.43) \times 10^8$
Activated sludge	6.4	94	n.d.[c]	$(14.90 \pm 0.20) \times 10^8$
Compost	7.9	54	n.d.[c]	$(6.18 \pm 0.53) \times 10^8$
Sand soil	8.5	8	1.3	$(2.43 \pm 0.11) \times 10^8$
Loam soil	8.2	21	2.4	$(11.10 \pm 0.08) \times 10^8$

[a] The percentage of organic matter content refers to dry sample material.

[b] Prokaryotic cell numbers (mean count ± standard deviation, n = 3) were determined by a DTAF-based direct microscopic cell count procedure (Bloem et al., 1994).

[c] Not determined (n.d.) as the organic matter content of dried sample material is obviously close to 100 %.

3.2. Co-extraction of eukaryotic DNA

The portions of bacterial, archaeal, and eukaryotic nucleic acids in the recovered DNA extracts were determined by consecutive hybridization of environmental DNA with three domain-specific oligonucleotide probes (Table 1). To calculate absolute amounts of DNA from hybridization signals, response factors (hybridization signal/ng of DNA) were determined for each probe by using known amounts of genomic control DNA (*Bacillus megaterium* representing Bacteria, *Pyrococcus furiosus* for Archaea, and *Hansenula polymorpha* for Eukarya). Hybridization assays with these control DNAs confirmed the high specificity of the probes since no cross-hybridization was detected. As an additional control, a fourth probe that binds to all three types of DNA was used. The responses obtained with this universal probe (Table 1) were in good agreement with the responses that were expected on basis of the composition of DNA extracts, which were found with the domain-specific probes.

DNA quantification of complex communities by hybridization with probes that target small subunit rDNA suffers from the fact that the number of probe targets, i.e. the number of rRNA operons, varies between organisms. While archaeal cells typically contain 1 or 2 *rrn* copies, bacteria have been found to comprise about 4 copies in average (Ribosomal RNA Operon Copy number Database, http://rrndb.cme.msu.edu; Klappenbach et al., 2000). Regarding the copy numbers of the control organisms *P. furiosus* (1 copy) and *B. megaterium* (7 copies), quantification of prokaryotic organisms should maximally deviate by a factor 2. Concerning eukaryotes, four-fold variation in rDNA copy number has been found for yeasts (Maleszka and Clark-Walker, 1993), but a comprehensive study of copy number in higher eukaryotes is still lacking. The assumption that targeted species on average have the same *rrn* copy numbers as the control organisms should therefore be regarded as an approximation, especially in the case of eukaryotes. However, total amounts of DNA determined by hybridization experiments agreed reasonably well with those found by gel electrophoresis, which excludes a major quantification bias.

Table 3. Lysis and cell extraction efficiencies of prokaryotic organisms and total DNA yields

	Direct DNA extraction			
	Soft lysis		Harsh lysis	
Sample	LE[a] [%]	µg g^{-1} sample[b]	LE[a] [%]	µg g^{-1} sample[b]
Marine sludge	91	10.9 ± 0.9	92	8.2 ± 0.1
Activated sludge	98	71.4 ± 2.3	99	107.8 ± 17.0
Compost	86	15.3 ± 1.5	89	7.0 ± 1.0
Sand soil	96	11.5 ± 1.8	99	9.6 ± 0.7
Loam soil	97	42.6 ± 4.6	99	36.5 ± 3.2

	Indirect DNA extraction			
	Blending		Cation-exchange	
Sample	CEE[c] [%]	µg g^{-1} sample[b]	CEE[c] [%]	µg g^{-1} sample[b]
Marine sludge	24	0.7 ± 0.0	7	0.1 ± 0.0
Activated sludge	68	5.0 ± 0.0	19	1.3 ± 0.1
Compost	39	1.2 ± 0.1	19	0.5 ± 0.0
Sand soil	79	0.8 ± 0.0	21	0.1 ± 0.0
Loam soil	55	3.3 ± 0.3	14	0.8 ± 0.1

[a] LE, lysis efficiency of indigenous prokaryotic organisms (mean value, n = 3) reached by direct DNA extraction methods. Coefficients of variation were below 7 % for all data.

[b] Total DNA yields (mean value ± standard deviation, n = 3 for direct lysis protocols, n = 2 for indirect methods). Values are related to g (wet weight) of environmental sample material.

[c] CEE, cell extraction efficiency (mean value, n = 2) reached by indirect DNA extraction methods. Lysis efficiencies for the extracted cells were larger than 99 % as determined by microscopical means. Coefficients of variation were below 7 % for all data.

Depending on the sample material used, direct DNA extracts contained 61 % to 93 % of eukaryotic nucleic acids (Fig. 2), which may be due to the partial lysis of indigenous eukaryotic organisms such as fungi, algae, and protozoa, or it may be caused by lysis of residual plant material. This is not surprising, as conditions similar to those of the harsh lysis protocol have been described for the lysis of yeasts, fungi, and plant and animal tissues (Borneman et al., 1996; Möller et al., 1992; Rose et al., 1990; Tebbe and Vahjen, 1993). Plant DNA can be even accessed with simple hot SDS treatment after initial disruption of the tissue structure by freezing and grinding (Edwards et al., 1991). Nevertheless, the observed large extent of co-extraction of eukaryotic DNA was not expected and shows that even though they yield a lot of

DNA, direct lysis methods should be avoided when gene banks are prepared in bacterial hosts strains. In contrast, DNA obtained by cell extraction was primarily derived from bacterial cells (> 92 %) due to the separation from eukarya by differential centrifugation, which makes it suitable for expression cloning. Only from marine sludge, large amounts of eukaryotic nucleic acids were also obtained after cell extraction, which may be due to the high content of easily detachable microalgae in this environment.

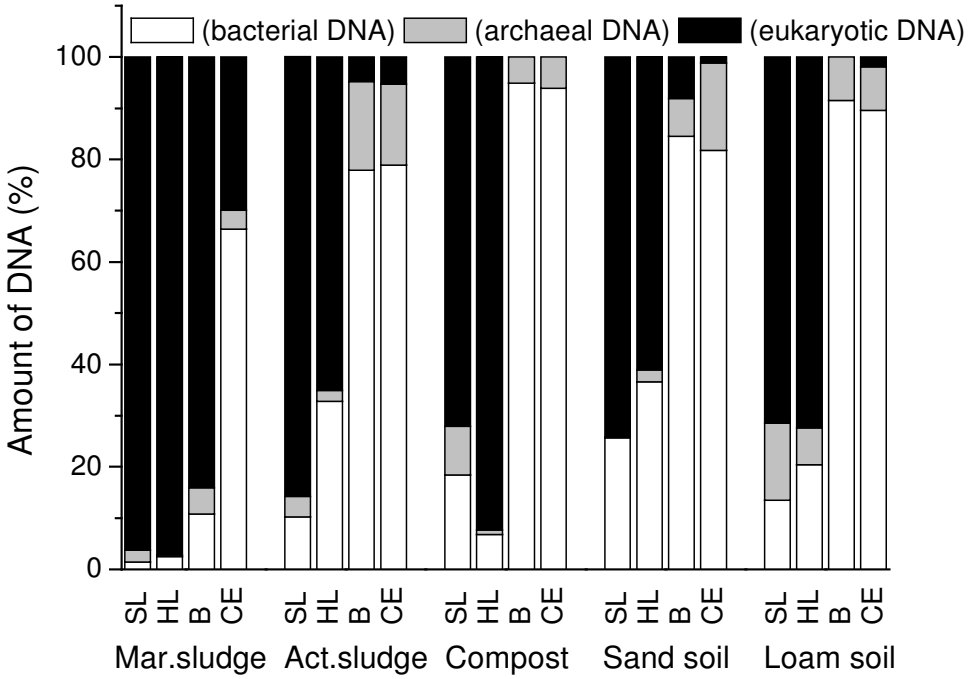


Figure 2. Origin of DNA prepared by soft lysis (SL), harsh lysis (HL), blending (B), and cation-exchange (CE) protocols. White columns correspond to bacterial DNA, gray columns to archaeal DNA, and black columns to eukaryotic DNA.

3.3. Bacterial diversity

PCR-DGGE is a widely used technique for profiling microbial communities in a variety of ecosystems (e.g. Heuer and Smalla, 1997; Nübel et al., 1999). A drawback of the method, however, is its rather high detection limit that only allows revelation of target genomes accounting for more than 0.1 % (Gelsomino et al., 1999) or even 1 % (Casamayor et al., 2000) of the total population. Despite this rather low sensitivity, PCR-DGGE analysis can provide valuable information about the redundancy that is to be expected in an environmental gene bank of limited size by revealing the number and frequency of the most abundant species in DNA extracts. Obviously, for the construction of gene banks with maximal biotransformation diversity, a high complexity of the source DNA is of advantage. Typical DGGE profiles of environmental DNA obtained by soft lysis (A) and blending (B) are shown in Fig. 3. When analyzing eight intense bands of activated sludge DNA prepared by blending, only one band (no. 6) contained more than one dominant 16S rDNA sequence (Table

4). Sequences fell into at least six different genera, including Gram-positive and Gram-negative species, indicating that a broad spectrum of bacteria was accessed by the blending protocol.

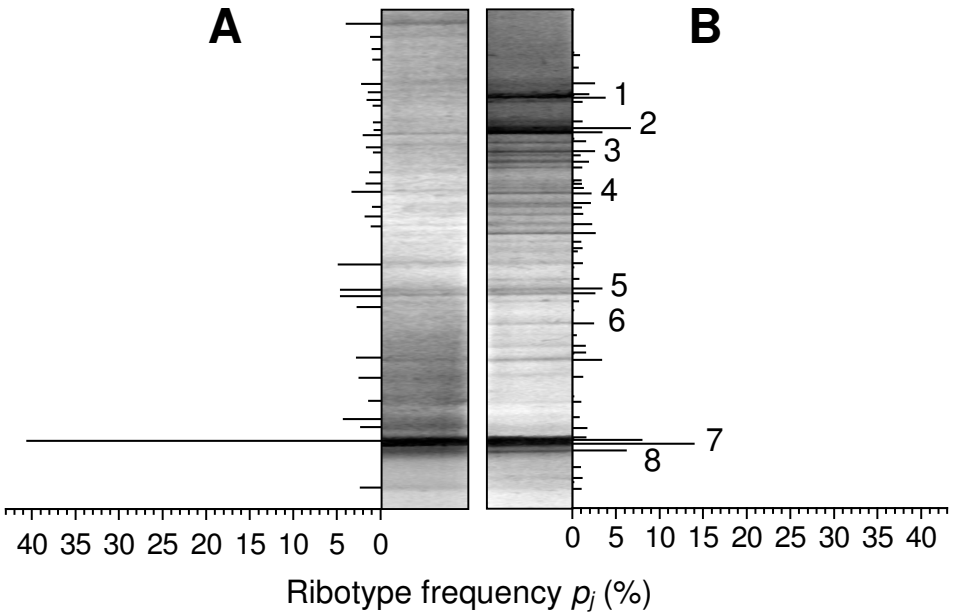


Figure 3. DGGE profiles of DNA extracted from activated sludge by soft lysis (**A**) and blending (**B**). Pixel density plots of the digitized images were used to determine the exact position, number (richness *S*) and relative abundance (p_j) of bands. Ribosomal RNA gene fragments present in the labeled bands (numbers 1 to 8) were sequenced and assigned to their closest database relatives (Table 4). The DGGE profiles are 47 % identical as calculated by 100×(2×common bands)/(total bands).

The complexity of DGGE profiles can be described by the number of bands present (richness *S*) and the Shannon-Weaver diversity index *H'*, which incorporates both the number of bands as well as their intensities (relative abundance p_j) (Pielou, 1977). While counting of bands is only limited by the resolution of the gel and the minimal amount of DNA required for detection, determination of relative band abundance might be compromised by biased PCR amplification. As a measure to prevent PCR bias, we performed reactions with dilution series of template, ensuring that the overall amplification process was still in its exponential phase (not shown). Other potential sources of bias could not be excluded, however. Varying G/C contents of target DNA, primer degeneracies, or even the absence of primer binding sites in yet unknown organisms have been suspected to cause differential amplification (von Wintzingerode et al., 1997). To minimize these effects, we chose a primer pair (U968-GC and L1401) that has been shown to equally amplify an approximately 400-bp16S rDNA segment from a broad range of bacterial taxa and uncultured soil bacteria (Engelen et al., 1995). Integrated amplification efficiencies were found to be similar for all DNA extracts used, which shows that compounds interfering with PCR amplification had been equally removed from all samples after purification on agarose gels.

Table 4. Phylogenetic origin of 16S rRNA gene segments PCR amplified from activated sludge DNA

DGGE band[a]	Closest RDP/GenBank relative		Similarity [%]
	Phylogenetic group	Genus and species, accession number	
1	Gram-positive bacteria	*Butyrivibrio fibrisolvens*, U77339	95
2	β-Proteobacteria	*Acidovorax* sp. 7078, AF078767	99
3	β-Proteobacteria	*Hydrogenophilus thermoluteolus*, AB009829	99
4	β-Proteobacteria	*Pseudomonas saccharophila*, AF368755	99
5	Gram-positive bacteria	unidentified bacterium from activated sludge, Z94008	97
6	Gram-positive bacteria	*Staphylococcus epidermidis*[b, c], AY030342	99
7	Gram-positive bacteria	*Lactosphaera pasteurii*, *L76599*	99
8	Gram-positive bacteria	*Staphylococcus epidermidis*[c], AY030342	99

[a] DGGE band numbers refer to the numbering of bands in Fig. 3 (B).
[b] Band no. 6 contained at least two non-identical sequences (99 % similarity) both aligning to the same RDP/GenBank entries.
[c] The 16S rDNA segments found in DGGE band 6 and 8 were not identical (99 % similarity).

Although it is commonly assumed that higher DNA yield equates to larger diversity, DNA extracts prepared by indirect methods were in most cases more diverse than those obtained by direct lysis regarding both diversity indices (Table 5). This seems surprising, because significantly lower proportions of indigenous microorganisms were accessed by cell extraction procedures (CEE versus LE, Table 3), which is thought to be the result of selective recovery of easily detachable cells. Nevertheless, very different types of bacteria were recovered, as shown for activated sludge (Table 4). The higher diversity of DNA found with indirect methods may be explained by the presence of abundant bacterial species that cannot be efficiently recovered by cell extraction due to shielding by the sample matrix, but that are readily accessed by direct lysis agents. Due to the predominance of these species, DNA extracts prepared by direct methods should be of lower complexity than those obtained by indirect procedures. Furthermore, the initial composition of free DNA released by direct lysis may be altered. Effects like selective and irreversible adsorption of DNA to the sample matrix, chemical DNA degradation, or shearing of less stable DNA molecules due to relatively vigorous extraction conditions can in principle lead to the loss of genetic information especially of Gram-negative species. At present, it remains difficult to judge which approach provides the better analysis of bacterial communities in terms of quantitative composition. For the construction of gene banks, however, this is not a major issue as this technique aims at qualitative results, i.e. revealing the presence of certain genes in a bacterial community. To maximize the chance of cloning all genes present in a microbial community, richness *S* and Shannon-Weaver index *H'* describing a DNA extract should both be as large as possible.

Besides ribotype richness and diversity, also the composition of DGGE profiles was strongly affected by the DNA recovery method used (Fig. 3). When comparing the similarity of DGGE profiles obtained by the two basic approaches, differences including up to 80 % of the detected bands were observed. Even within the same category of protocols, DGGE profiles varied at least by 20 %, emphasizing the impact of the DNA recovery method chosen on the outcome of any ecological experiment.

Table 5. Diversity indices based on 16S rRNA gene segments amplified from environmental DNA

Sample	Direct DNA extraction				Indirect DNA extraction			
	Soft lysis		Harsh lysis		Blending		Cation-exchange	
	H' [a]	S [b]	H' [a]	S [b]	H' [a]	S [b]	H' [a]	S [b]
Marine sludge	3.1	16	3.0	11	4.1	25	4.0	21
Activated sludge	3.7	30	4.0	31	5.0	47	4.3	37
Compost	4.5	47	4.3	35	4.3	32	4.6	29
Sand soil	4.5	28	4.2	29	4.4	42	4.2	31
Loam soil	4.7	32	4.6	31	5.2	44	5.1	43

[a] Shannon-Weaver indices H' were calculated by $H' = -\sum p_j \log_2 p_j$, where p_j is the relative abundance of the j^{th} band in a DGGE profile.
[b] Richness S corresponds to the total number of distinct bands in a DGGE profile.

3.4. Suitability of DNA for cloning

To date, most gene banks described in the literature have been constructed in cosmid, fosmid, lambda or plasmid expression vectors using relatively small insert sizes (Cottrell et al., 1999; Henne et al., 2000; Majerník et al., 2001; Vergin et al., 1998). In terms of molecular weight, DNA extracted by soft lysis in this study was suitable for cloning in these systems, making insert sizes of up to 30 kb possible (Fig. 4). By omitting the separation of soil particles before chloroform extraction as well as the re-extraction of soil pellets, Rondon et al. (2000) succeeded in the construction of a bacterial artificial chromosome (BAC) library with an average insert size of even 44.5 kb. When we included bead beating in the procedure, DNA was strongly sheared, which may compromise the cloning of gene-sized fragments (Fig. 4). Harsh lysis conditions comparable to those described in the present study should therefore be avoided if DNA is used for expression cloning. Cell extraction methods, in contrast, routinely recovered DNA of high molecular weight (> 50 kb), allowing the preparation of large insert gene banks.

To compare the purity of DNA extracts obtained by both DNA extraction strategies, DNA recovered from three environmental samples by soft lysis and blending, respectively, was used for blunt-end cloning in pZero-2, a high-copy plasmid vector. Inserts were prepared by mechanical shearing, which avoids selective, methylation pattern-dependent fragmentation of DNA and the possible loss of whole genomes for cloning as encountered when cutting DNA by enzymatic digestion.

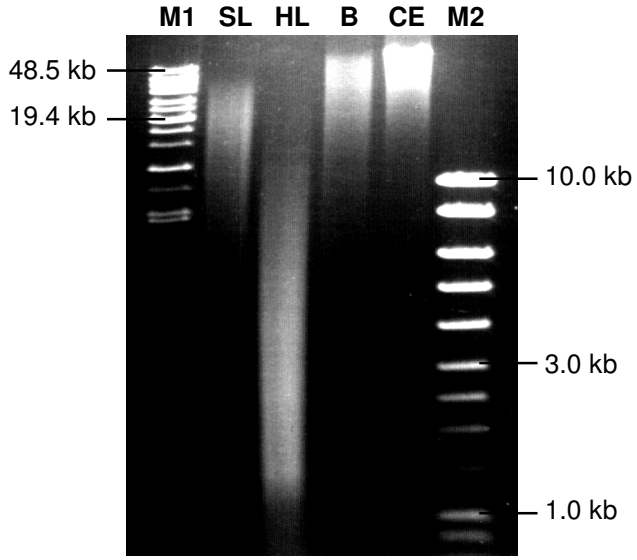


Figure 4. Molecular weight of DNA fragments isolated by soft lysis (SL), harsh lysis (HL), blending (B), and cation-exchange (CE) protocols from loam soil. Fragment size distributions obtained by the different methods are representative for all environmental samples studied. Lane M1: High Molecular Weight DNA (Invitrogen). Lane M2: Smart Ladder molecular weight marker (Eurogentec).

Between two- and three-fold higher cloning efficiencies, i.e. transformants per µg of insert DNA, were obtained with nucleic acids isolated by the blending method as compared to the soft lysis protocol (Table 6). When cells were extracted with Chelex 100, even higher efficiencies were reached, as we found with activated sludge (data not shown). Due to its low total DNA yield, however, we recommend the use of this method only when purity is of paramount concern or levels of indigenous eukaryotic microorganisms are extremely high, such as e.g. in marine sludge. Worst results were found with DNA recovered from compost, probably due to co-extracted polyphenolic substances like humic acids that are known to interfere with various enzymatic reactions and decrease transformation efficiencies (Tebbe and Vahjen, 1993). Besides with activated sludge, high cloning efficiencies were reached with DNA extracted from loam soil with the blending protocol.

Table 6. Cloning efficiencies of environmental DNA

Sample	Soft lysis [10^3 transformants μg^{-1} DNA][a]	Blending [10^3 transformants μg^{-1} DNA][a]
Activated sludge	52.9 ± 5.5	107.8 ± 13.9
Compost	3.2 ± 1.0	9.6 ± 0.7
Loam soil	17.7 ± 2.8	44.4 ± 18.3

[a] Cloning efficiencies (mean value ± standard deviation, n = 3) are given as number of transformants obtained per µg of environmental DNA fragments used for ligation in pZero-2.

To estimate the functional diversity present in the loam soil gene bank, we conducted initial screens for five different enzymatic activities on agar media. When screening 80,000 clones that comprised about 200 Mb of total insert DNA as determined by enzyme restriction analysis, two clones expressing amidase activity, 3 clones degrading β-D-glucovanillin, 4 clones with β-lactamase activity, and 1 clone with amylase activity were found. Lipase activity was not revealed. To assure plasmid encoded activity, recombinant plasmids were isolated from the identified positive clones and retransformed to *E. coli* TOP10 cells. Although the gene bank was not sufficiently large to cover the complete genomes of the at least 44 abundant species present in the sample (Table 5), clones expressing 4 from the 5 enzymatic activities tested could be isolated. This shows that an environmental gene bank of high quality can be prepared from DNA extracted by the blending protocol.

Although the indirect DNA extraction protocols recovered smaller amounts of total DNA than the direct methods, similar numbers of clones carrying prokaryotic DNA should be obtained by either method due to the higher selectivity of the indirect approach towards prokaryotic DNA and higher cloning efficiencies. The apparent drawback of the indirect approach, the low DNA yields, may be overcome by adjusting the experimental conditions to the specific environmental material used. A study of Duarte et al. (1997) suggests, for instance, that DNA yields close to the theoretical maximum can be obtained from different soils with a cell extraction method based on shaking with gravel in pyrophosphate buffer. Although the four protocols described in this paper have not been optimized for the largely differing environmental samples studied, a clear trend towards higher bacterial diversity, lower eukaryotic DNA content, and superior purity has been observed for DNA isolated by indirect means. Our results therefore indicate that the use of indirect DNA isolation strategies generally results in environmental gene banks of better quality.

ACKNOWLEDGEMENTS

We thank M. van der Velde and S. den Dulk from Bioclear (Groningen, The Netherlands) for DGGE analysis.

3

Quantifying the accessibility of the metagenome by random expression cloning techniques

Esther M. Gabor, Wynand B. L. Alkema, and Dick B. Janssen

Accepted for publication in *Environmental Microbiology*

The exploitation of the metagenome for novel biocatalysts by functional screening is determined by the ability to express the respective genes in a surrogate host. The probability of recovering a certain gene thereby depends on its abundance in the environmental DNA used for library construction, the chosen insert size, the length of the target gene, and the presence of expression signals that are functional in the host organism. In this paper, we present a set of formulas that describe the chance of isolating a gene by random expression cloning, taking into account the three different modes of heterologous gene expression: independent expression, expression as a transcriptional fusion, and expression as a translational fusion. Genes of the last category are shown to be virtually inaccessible by shotgun cloning because of the low frequency of functional constructs. To evaluate which part of the metagenome might in this way evade exploitation, 32 complete genome sequences of prokaryotic organisms were analyzed for the presence of expression signals functional in *E. coli* hosts, using bioinformatics tools. Our study reveals significant differences in the predicted expression modes between distinct taxonomic groups of organisms and suggests that about 40 % of the enzymatic activities may be readily recovered by random cloning in *E. coli*.

1. INTRODUCTION

During the past five years, random cloning of microbial DNA directly isolated from environmental materials and subsequent screening of expression libraries for the presence of a desired enzyme activity has become a useful tool for the discovery of novel biocatalysts. The collective genomes of microbes indigenous to a certain habitat, now often referred to as the metagenome (Handelsman et al., 1998), are considered to be an almost inexhaustible source of new enzymes (Cowan, 2000). Indeed, screening of the metagenome has already yielded various new biocatalysts (for a recent review see Lorenz and Schleper, 2002), and with steadily improving techniques this number is expected to rise quickly. In most cases, gene banks are screened with activity-based assays as they allow the recovery of completely new types of enzymes without any prior knowledge of the sequence, relying only on the ingenuity of the screening method. Such a functional screening, however, requires gene expression and proper folding of the resulting protein in a heterologous host, most frequently *E. coli*, which is not always easily achieved.

The minimal set of requirements for gene expression includes the presence of a promoter for transcription, and a ribosome binding site (rbs) in the –20 to –1 region upstream of the start codon for initiation of translation. Both sites must be suitable for the expression machinery of the bacterial host cell. Besides these *cis*-acting DNA sequences, the formation of an active protein may also rely on *trans* factors that need to be provided by the host organism such as special transcription factors, inducers, chaperones, cofactors, protein-modifying enzymes, or a proper secretion machinery. Whether or not essential *trans* factors are present in the host is in most cases difficult

or even impossible to predict. In contrast, functional *cis* elements can be identified based on DNA sequence analysis (e.g. Ermolaeva et al., 2000; Gold et al., 1981; Staden, 1983).

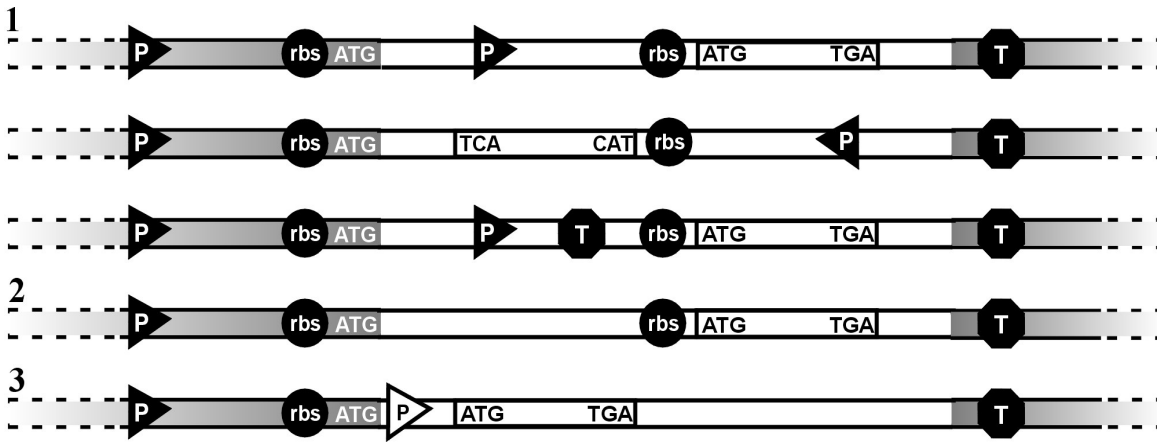


Figure 1. Different modes of gene expression in a heterologous bacterial host. Vector DNA is shaded in gray, insert DNA is white. P, promoter; T, transcription terminator. (1) Independent gene expression (IND). The insert can be cloned in either direction. However, if a transcriptional termination sequence is located in-between the present promoter and the cds start (1, bottom), the gene can only be expressed as a transcriptional fusion. (2) Expression based on a transcriptional fusion (TRANSC). (3) Expression dependent on expression signals located on the vector (DEP). Here, a promoter that is active in the host system may be present (white triangle). However, a suitable rbs is lacking, which is why the gene can only be expressed as a fusion protein.

To assure the formation of mRNA transcripts of heterologous coding sequences (cds), vectors carrying their own strong promoter (and possibly a transcriptional terminator) are usually employed in expression cloning. In addition, a rbs followed by a bacterial start codon in favorable spacing (9 bp for *E. coli*) is generally supplied close to the multiple cloning site. In such systems, three modes of gene expression can be anticipated: (1) independent gene expression with both the promoter and the rbs provided by the insert (IND), (2) expression as a transcriptional fusion with only the rbs located on the insert (TRANSC), and (3) expression as a translational fusion depending on both the promoter and the rbs of the vector (DEP) (Fig. 1). Intuitively, it can be understood that the occurrence of a functional translational fusion is very rare and, consequently, the chance of discovering a microbial gene devoid of expression signals that are recognized by its heterologous host is low. Transcriptional fusions, in contrast, are more likely to occur, requiring only that the gene of interest is cloned in the correct orientation and is not separated from the plasmid-localized promoter by a transcriptional termination sequence. If all expression signals assigned to a certain gene are recognizable by the host strain, expression is independent from vector signals and identification of the gene due to the activity of its product is most likely, provided that no repression occurs.

Many studies have revealed the enormous diversity of mostly unculturable bacterial species in different natural habitats (Bohannan and Hughes, 2003; Torsvik and Øvreås, 2002), which has served as an incentive to apply random expression cloning techniques to environmental DNA for the recovery of novel enzymatic activities. However, little is known about the question to which extent the metagenome can actually be exploited by this strategy, i.e. which fraction of the encoded enzymes can be expressed and, consequently, discovered in a heterologous host such as *E. coli*. In this study, we approached this question by deriving a set of formulas that can be used to calculate the number of clones that is needed for the comprehensive screening of an environmental DNA sample, taking into account the three basic types of heterologous gene expression. Furthermore, a bioinformatics approach was used to predict, in which mode the proteins encoded by the genomes of typical soil organisms are expressed in an *E. coli* host and, consequently, how readily they can be recovered by random expression cloning.

2. MATERIALS AND METHODS

The GeneClassifier program as outlined in the Results section was written as a pipeline of Perl scripts using Bioperl modules (Stajich et al., 2002). Positional frequency matrices (PFMs) for *E. coli* promoters and rbs were obtained from literature (Lisser and Margalit, 1993; Schneider and Stephens, 1990) and are shown in Fig. 2. Intrinsic transcription terminators were predicted with the TransTerm program (http://www.tigr.org/software/; Ermolaeva et al., 2000). The annotated sequence files of the genomes used in this report were downloaded from the GenBank database at the National Center for Biotechnology Information (NCBI).

3. RESULTS AND DISCUSSION

3.1 Statistical analysis

Whether or not a certain gene is discovered by random expression cloning statistically depends on the size of the screened gene bank. Even a gene without any expression signal can be expressed as a fusion protein and, consequently, be recovered, provided that a large enough gene bank (> 10^7 clones, see below) is tested. For practical reasons, however, the number of clones that can be prepared and screened is rather limited and typically ranges from 10^4 to 10^6 clones.

The number of clones required to detect an enzymatic activity with a certain probability depends on the expression mode of the corresponding gene. In principle, assaying a single clone that carries a random heterologous DNA fragment can be regarded as a *Bernoulli* experiment with only two possible outcomes: the respective clone exhibits activity or it is inactive. A positive result can be expected with the probability n_{active}/n_{total}, with n_{active} as the number of different active clones that can theoretically be constructed and n_{total} as the number of all possible constructs that can

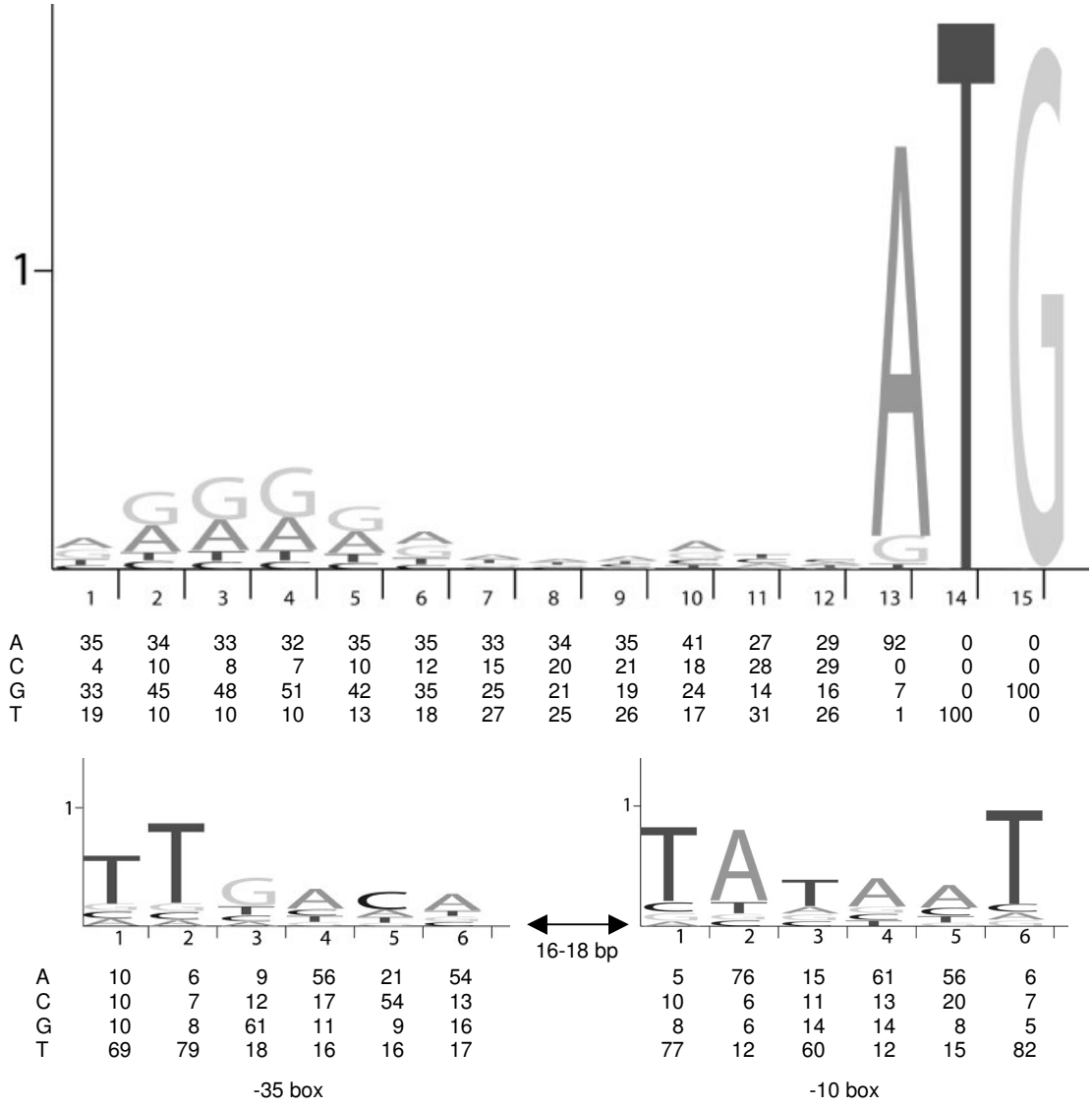


Figure 2. Sequence logos and positional frequency matrices (PFMs) of the *E. coli* rbs (top, Schneider and Stephens, 1990) and promoter (bottom, Lisser and Margalit, 1993) consensus sequences used in this study. Tabulated values are the frequencies (in %) of the four nucleotides A, C, G, T.

be made from a given DNA sample. As n_{total} is much larger than the typically screened library, n_{active}/n_{total} can be assumed to be constant for all tested clones. Consequently, the number of clones N_P that is required to recover a target gene at least once with the probability P, can be derived from a binomial distribution:

$$N_P = \frac{\ln(1-P)}{\ln(1-\frac{n_{active}}{n_{total}})} \qquad \text{Equation 1}$$

For gene banks prepared from a single organism, n_{active}/n_{total} is given by $(I\text{-}X)/(c \cdot G)$ as illustrated in Fig. 3, with I as the insert size, X as the size of the gene of interest, and G as the genome size. The correction factor c reflects the different possible expression modes, IND, TRANSC, and DEP as described below. For simplicity, the target enzyme is assumed to be encoded by a single copy gene.

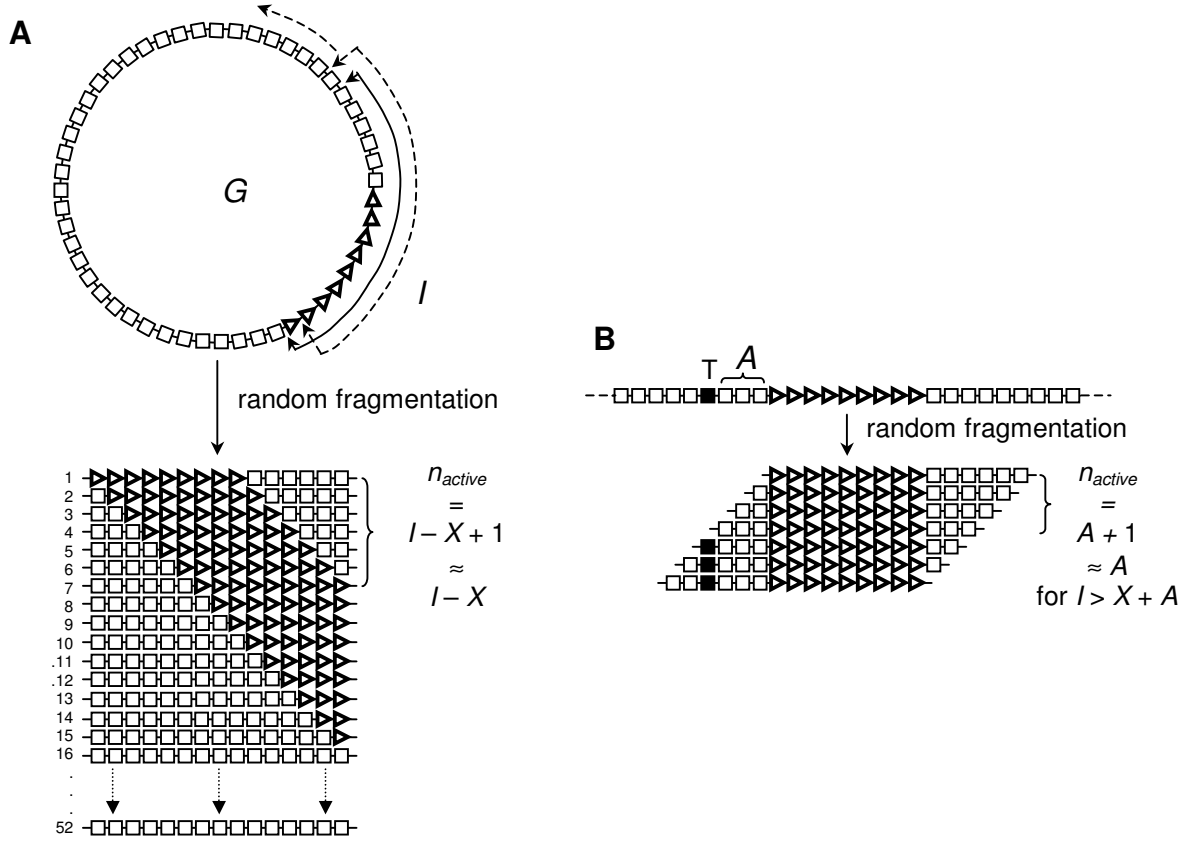


Figure 3. Schematic overview of the different inserts that can be produced by random fragmentation of a single genome, and exemplary calculation of n_{active} and n_{total}. To illustrate the principle, convenient but biologically meaningless values are chosen for the different parameters (genome size G = 52 bp, and insert size I = 15 bp). Only a single target gene (gene size X = 9 bp, including potential promoter and rbs sequences) is present (triangles). Inserts that are randomly prepared can start at any position in the genome, leading to a total of 52 (= G) different possible clones. **A.** In this way, $I\text{-}X$ inserts with a complete gene can be obtained, leading to a corresponding number of active clones in the case of independent gene expression. Consequently, n_{active}/n_{total} is $(I\text{-}X)/(G \cdot c)$ with c = 1. If transcriptional fusions are required, the inserts carrying a complete gene need to be cloned in the right direction behind the vector promoter, yielding active clones in only half of the cases as reflected by c = 2. **B.** If a transcriptional fusion is needed and $I > A+X$ with A as the distance between the gene start and the closest upstream terminator (T), only A active clones can be formed irrespective of the used insert size. In this case, n_{active}/n_{total} becomes $A/(2 \cdot G)$.

When libraries are prepared from metagenomic DNA, G becomes the average size of genome sequences present in the sample, and z is the number of different genomes (species) comprised, assuming their even distribution in the source DNA:

$$N_P = \frac{\ln(1-P)}{\ln(1-\frac{I-X}{G \cdot c \cdot z})} \qquad \text{Equation 2}$$

From all genome sequences of prokaryotes available to date at The Institute for Genomic Research (www.tigr.org), the average size is 3,100 kb, a number that can be used as a rough approximation for G. This value, however, is somewhat biased towards pathogenic organisms, which are predominant in the database and have been found to be considerably smaller than their free-living, environmental counterparts (Horn et al., 2003). An estimate for z could be obtained by analyzing different soil and sediment samples by denaturing gradient gel electrophoresis, revealing 25 to 44 different species (*Chapter 2*).

For independent gene expression, c is 1, leading to relatively small numbers of clones that need to be screened to recover a target gene (Fig. 3.A). If transcriptional fusions are required, however, c equals 2, since the genes need to be cloned in the correct orientation to the vector promoter and, consequently, N_P increases. For the DEP class of genes, N_P rises even further because here, $c = 6$ as genes must be specifically cloned into one of the 6 reading frames. It should be noted that these considerations for the TRANSC and DEP genes only hold true when relatively small inserts are used. For larger inserts, the situation becomes more complicated as described below.

It follows from Eq. 2 that by using larger insert sizes, N_P decreases, which is obviously a desired effect and especially holds for the IND fraction of genes. If transcription of a given cds, however, needs to be triggered from the vector molecule such as for members of the TRANSC group, also the chance of obtaining non-productive clones rises due to the increased chance of creating constructs that contain a transcription terminator upstream of the target gene. This fact causes N_P to become independent of the chosen insert size if $I > A + X$, A being the average distance between a start codon and its preceding transcription terminator (Fig. 3.B):

$$N_P = \frac{\ln(1-P)}{\ln(1-\frac{A}{2 \cdot G \cdot z})} \qquad \text{Equation 3}$$

Consequently, choosing inserts larger than $A + X$ will not further decrease the number of clones that need to be screened.

If translational fusions are required, the useful insert size is further reduced. Even if a cds is cloned into the right reading frame and no transcription terminator is located upstream of it, two situations can occur that compromise the formation of an active gene product: (1) a stop codon is located between the rbs/start codon stretch of the vector and the actual cds start, or (2) the amino acids fused to the N-terminus of the resulting fusion protein impede activity. The number of tolerated fused amino acids B

is difficult to determine in general terms although it can be anticipated that it is rather small, resulting in the requirement to use extremely large gene banks according to Eq. 4, which can be deduced in a similar way as Eq. 3.

$$N_P = \frac{\ln(1-P)}{\ln(1-\frac{3 \cdot B}{6 \cdot G \cdot z})} \qquad \text{for } I > 3 \cdot B + X \qquad \text{Equation 4}$$

3.2 *The GeneClassifier program*

As becomes apparent from the statistical considerations described above, the likelihood of discovering environmental genes, i.e. their accessibility by random cloning, is directly coupled to the way they can be expressed in a heterologous host. To estimate the fraction of the metagenome that can be mined using *E. coli* as the expression host, we developed a program called GeneClassifier (Fig. 4).

GeneClassifier locates putative promoter, rbs and transcriptional terminator sequences in complete genome sequences, using matrix searches with *E. coli* consensus sequences and the TransTerm program, respectively. The locations of cds within the genome sequences are taken from the annotations. Searches for promoter sequences are performed in the 2,500 bp preceding each cds, while rbs searches are restricted to the –12 to +3 region. The positions of possible promoters and rbs are determined by searching the genomic sequences with positional weight matrices (PWMs) (Stormo et al., 1986) that describe the consensus sequences for those sites and can be obtained from positional frequency matrix (PFMs). PFMs of *E. coli* promoters and rbs were taken from Lisser and Margalit (1993) and Schneider and Stephens (1990), respectively (Fig. 2). GeneClassifier scores sequences based on their match to the PWMs using the Perl TFBS modules (Lenhard and Wasserman, 2002). The absolute score of each site (S_{site}) is then converted to a relative score (S_{rel}), given by

$$S_{rel} = \frac{S_{site} - S_{\min}}{S_{\max} - S_{\min}} \qquad \text{Equation 5}$$

in which $S_{\min}$ and $S_{\max}$ are the scores of the worst and the best possible match to the PWM, respectively. All sites that score above a given cut-off value of S_{rel} are considered to be putative promoters or rbs and are assigned to the cds they are preceding. It should be noted that the values of the cut-offs reflect the respective promoter and rbs strength, and are thus directly related to the expression level. When using a sensitive screening assay, weaker promoter and rbs sites can in principle be accepted than when using tests that require high amounts of enzyme activity. The locations of intrinsic (rho-independent) transcription terminators are determined by screening the complete intergenic regions of the genomes with the TransTerm program (http://www.tigr.org/software/; Ermolaeva et al. 2000).

Based on the relative positions of the cds and the (possibly) present upstream expression signals, the transcriptional and translational context of each cds is then determined according to Fig. 1, and the frequencies of the three categories (IND,

TRANSC, and DEP) in the analyzed genome are calculated. Cds with a preceding rbs and a promoter, and without a transcription terminator in-between the promoter and the cds are classified as IND. Genes with only a predicted rbs as well as genes with a valid rbs and a promoter but with an intervening terminator are assigned to the TRANSC group of genes, while genes lacking both the rbs and the promoter are labeled as DEP.

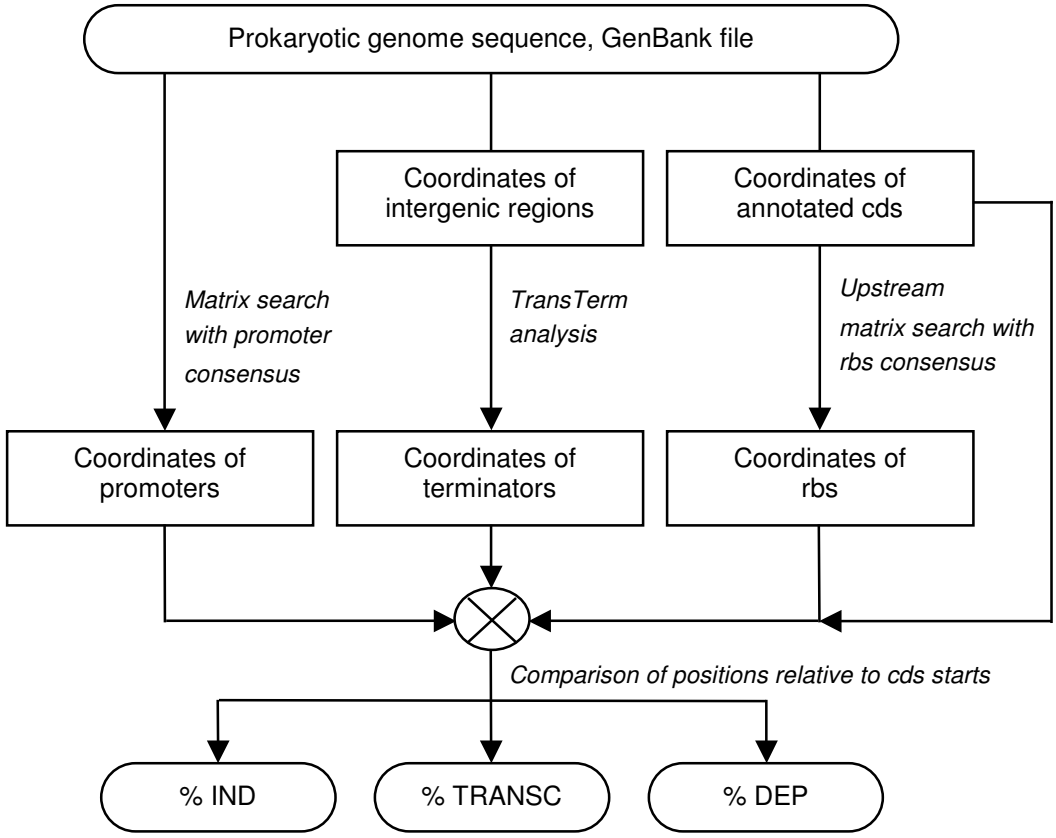


Figure 4. Architecture of the GeneClassifier program

3.3 Heterologous gene expression in E. coli

A total of 32 completely sequenced prokaryotic genomes (Table 1) were analyzed with the GeneClassifier program. As soil and sediment samples are mostly used in metagenomic cloning, we selected genomes of organisms that fall into different classes of bacteria known to be abundant in these habitats: Proteobacteria, Actinobacteria, and low-G/C Gram-positive Eubacteria, i.e. Firmicutes (Handelsman et al., 1998). Recently, evidence has been provided that also Archaea, organisms usually assigned to extreme environments, populate these habitats (Bintrim et al., 1997) and DNA samples extracted from different soil types and sediment were found to contain up to 17 % of archaeal genomic DNA (*Chapter 2*). Therefore, also 10 archaeal representatives were included in the expression analysis.

Table 1. Genomes analyzed in this study. Average distances between the closest terminator upstream of each cds and the start codon (*A*) were calculated with the GeneClassifier program.

Taxonomic group	Species (circular chromosome length, Mb)	Accession no.	GC-content (%)	*A* (kb)
Eury-archaeota	*Archaeoglobus fulgidus* (2.2)	AE000782	48.5	489
	Methanobacterium thermoautotrophicum (1.8)	AE000666	49.5	190
	Methanococcus jannaschii (1.7)	L77117	31.3	40
	Pyrococcus abyssi (1.8)	AL096836	44.6	354
	Thermoplasma acidophilum (1.6)	AL139299	45.9	111
	Methanopyrus kandleri (1.7)	AE009439	61.1	118
	Methanosarcina mazei (4.1)	AE008384	41.4	36
Cren-archaeota	*Aeropyrum pernix* (1.7)	BA000002	56.2	81
	Pyrobaculum aerophilum (2.2)	AE009441	51.3	633
	Sulfolobus solfataricus (3.0)	AE006641	35.7	287
Firmicutes (low-G/C Gram-positive)	*Clostridium acetobutylicum* (3.9)	AE001437	30.9	5
	Listeria monocytogenes (3.0)	AL591824	37.3	4
	Staphylococcus aureus (2.8)	BA000018	32.7	5
	Streptococcus pneumoniae (2.0)	AE007317	39.6	4
	Bacillus subtilis (4.2)	AL009126	43.5	4
Actino-bacteria (high-G/C Gram-pos.)	*Streptomyces coelicolor* (8.7)	AL645882	72.1	49
	Bifidobacterium longum (2.2)	AE014295	60.0	7
	Corynebacterium efficiens (3.1)	BA000035	53.7	11
	Mycobacterium leprae (3.3)	AL450380	57.7	267
Proteo-bacteria (Gram-negative)	*Brucella melitensis* (2.1)	AE008917/8	57.0	11
	Rickettsia prowazekii (1.3)	AJ235269	29.1	155
	Agrobacterium tumefaciens (2.8)	AE007869	59.3	8
	Caulobacter crescentus (4.0)	AE005673	67.1	11
	Neisseria meningitidis (2.3)	AE002098	51.4	3
	Escherichia coli (4.6)	U00096	50.7	8
	Pseudomonas aeruginosa (6.3)	AE004091	66.4	16
	Yersinia pestis (4.7)	AL590842	47.5	8
	Haemophilus influenzae (1.8)	L42023	38.0	4
	Vibrio cholerae (3.0)	AE003852/3	47.6	5
	Xylella fastidiosa (2.7)	AE003849	52.6	54
	Heliobacter pylori (1.7)	AE000511	38.8	65

The choice of the cut-off values for the promoter and rbs searches (see above) can have a significant impact on the outcome of the GeneClassifier analysis. Results for the 5 taxonomic groups of organisms at different combinations of threshold settings are shown in Fig. 5. It turned out that the fraction of genes in the IND class decreases with increasing cut-off values for both the rbs and the promoter sites, whereas the fraction of genes in the TRANSCR class is mainly determined by the respective promoter cut-off. The curves for the different phyla, however, do not cross each other, except in the extreme high-rbs/high-promoter cut-off region of the TRANSC category, which indicates that the differences between taxonomic groups are found for a wide range of biologically meaningful parameter settings.

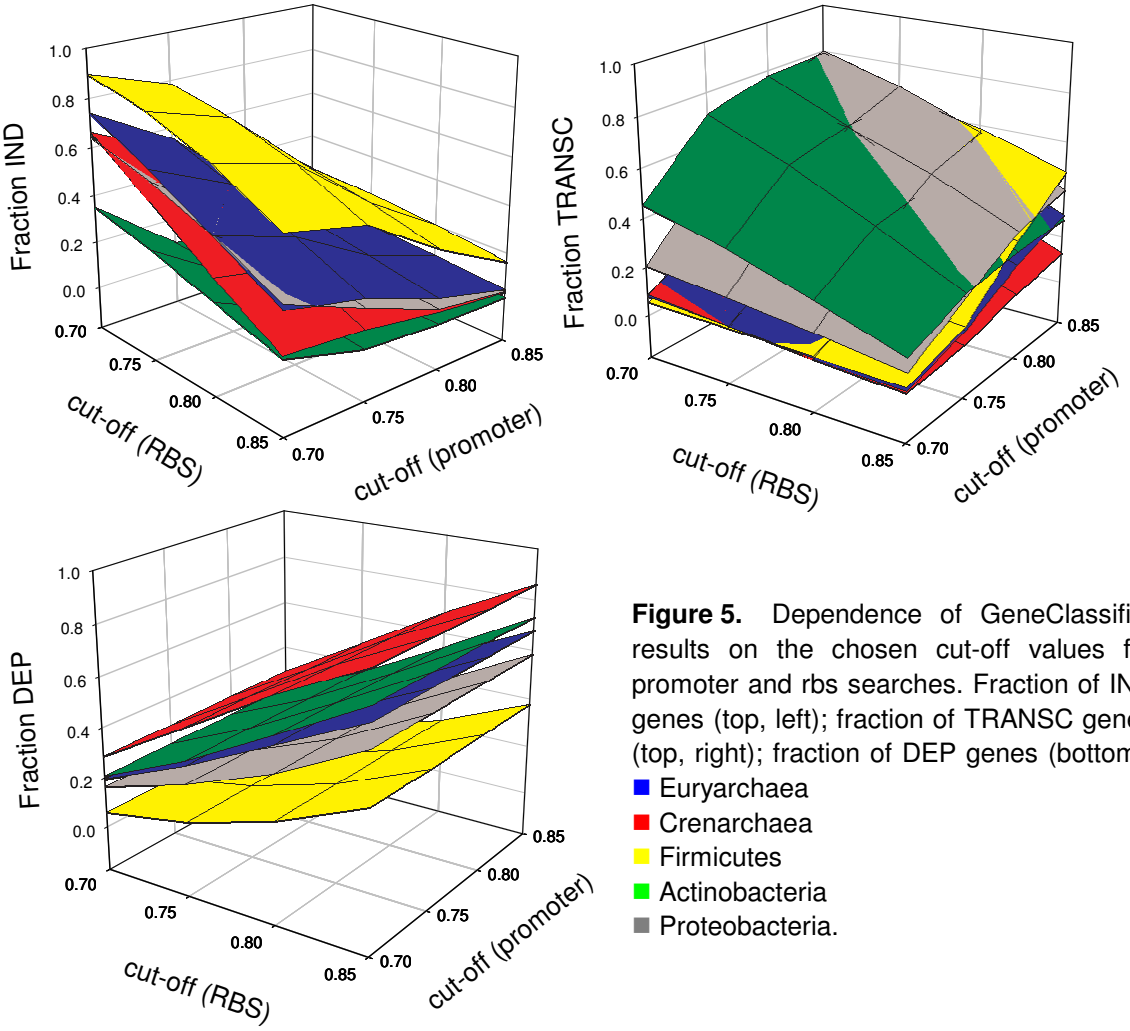


Figure 5. Dependence of GeneClassifier results on the chosen cut-off values for promoter and rbs searches. Fraction of IND genes (top, left); fraction of TRANSC genes (top, right); fraction of DEP genes (bottom).
- Euryarchaea
- Crenarchaea
- Firmicutes
- Actinobacteria
- Proteobacteria.

For the sake of clearness, GeneClassifier results are shown for a single cut-off combination (0.75 for both the promoter and the rbs) in Fig. 6. Firmicutes were predicted to have the largest fraction of independently expressible genes (73 %), which is agreement with experimental observations of Handelsman et al. (1998) who found that more than 50 % of the traits of *Bacillus cereus* that they tested were readily expressed in *E. coli*, presumably from their own expression signals. Actinobacteria, in

contrast, only contained 7 % of IND genes. Surprisingly, the phylum of Proteobacteria, containing *E. coli* itself, did not constitute the group with the largest IND fraction. This alleged paradox can be partially explained by the different GC-contents of the analyzed taxonomic groups (Table 1). Firmicutes have a relatively low GC-content compared to Actinobacteria (37 % versus 61 % in average), which statistically leads to a higher chance of matches to the AT-rich promoter consensus in the first group of organisms. Corresponding to their GC-contents of 47 and 51 %, respectively, Archaea and Proteobacteria ranked at an intermediate position with respect to their IND genes. The apparent presence of many promoters in Archaea, however, is not only explained by statistical reasons. It is known that archaeal promoter sequences resemble eukaryotic promoters in structure and function, but also contain a TTTAWATA (W = A or T) motif about 20 bp upstream of a less important but AT-rich initiator element (Brown et al., 1989). These two sequence stretches closely match the *E. coli* –10 and –35 promoter boxes, respectively.

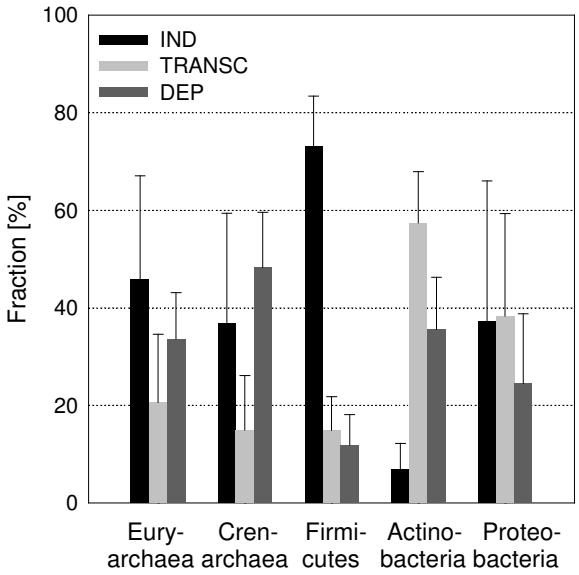


Figure 6. GeneClassifier results for the different taxonomic groups analyzed. For both promoter and rbs searches a threshold of 0.75 was chosen.

Interestingly, average distances between cds starts and preceding transcription terminators, *A*, were very large for all archaeal genomes analyzed by GeneClassifier, ranging clearly above 100 kb for most genomes (Table 1). Apparently, intrinsic transcription termination is not a common mechanism in this group of organisms, which is in agreement with the findings of others (Washio et al., 1998). For expression cloning of the TRANSC fraction of archaeal genes, the absence of transcription terminators is, in fact, a convenient feature since there is no disturbance of transcriptional fusions, and large insert sizes can be used to effectively decrease the number of clones that needs to be screened (see previous section).

In eubacterial genomes, in contrast, *A* was found to be 15 kb in average, excluding the *M. leprae* and *R. prowazekii* genomes, for which unusually low numbers of

transcription terminators were predicted. As can be seen in Fig. 6, the TRANSC category constitutes the largest fraction of genes of Actinobacteria and Proteobacteria, and still comprised 15 % of the genes of Firmicutes. For the cloning of this large number of genes, a small insert size of around 15 kb thus appears optimal, taking into account the increasing experimental expenditure when working with large DNA fragments and the higher expression levels that can usually be reached in high copy number vectors that are suitable for small insert sizes. It is obvious that these considerations only hold when targeting single genes or small operons. Especially if screening attempts to discover novel natural products, which often require the expression of large biosynthetic gene clusters, large insert vectors such as bacterial artificial chromosomes (BACs) remain the most appropriate vector systems (MacNeil et al., 2001).

With 34 % to 48 %, Archaea and Actinobacteria contained the largest fractions of genes depending on a translational fusion to the vector molecule and, therefore, most of their DNA is virtually inaccessible by conventional random expression cloning. As expected, genes of Firmicutes and Proteobacteria were identified to be most readily expressed in *E. coli*.

4. CONCLUSION

This study constitutes a quantitative approach to the question to which extent the metagenome can be exploited by the current random expression cloning techniques. The probability of detecting a certain enzyme activity is directly correlated with the expression mode of the respective cds in a heterologous host. As summarized in Fig. 7, genes that are preceded by expression signals that are functional in *E. coli* can be recovered by screening a relatively small number of clones, a number that exponentially decreases when using larger insert sizes. About 40 % of the genes of all genomes analyzed in this study were predicted to be readily expressible in this way, with strong variations between different groups of organisms (7 to 73 %).

The expression of the majority of genes, in contrast, was found to be dependent on expression signals located on the cloning vector. One third requires transcription to be triggered from the vector promoter. To recover genes from the TRANSC category, significantly bigger gene banks need to be constructed than for the IND fraction, particularly when working with large inserts. In fact, using inserts > 15 kb does not seem to be useful given the abundance of transcription terminators that may interfere with the formation of a complete transcript. Another 30 % of all analyzed genes, again with strong deviations between different phyla, can only be expressed as fusion proteins in *E. coli* due to the lack of suitable expression signals. This demands for gene libraries that comprise more than 10 millions of clones irrespective of the used insert size (Eq. 4, Fig. 7), which is not feasible in the scope of most projects. Consequently, genes falling into the DEP category can be regarded to be virtually inaccessible by random expression cloning.

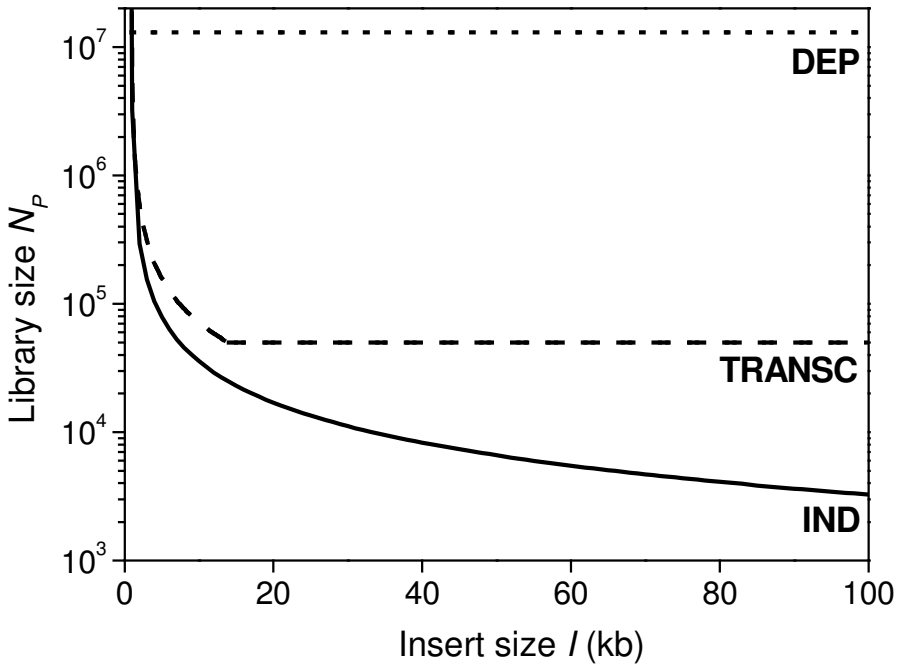


Figure 7. Dependence of the library size on the chosen insert size for the three different expression modes. Curves were calculated with Eq. 2-4 and P = 0.9, X = 0.9, G = 3,200 kb, z = 44, A = 13 kb, and B = 50.

It should be noted that the here-presented estimates of genes that can be detected in functional screenings are relatively optimistic, constituting an upper limit rather than an absolute number, since enzyme activity does not only require protein expression, but also proper folding, possibly incorporation into the cell membrane, secretion, or the presence of specific cofactors and modifying enzymes. Compared to expression mechanisms, post-translational processes are though much more complex and diverse, making predictions on their occurrence in heterologous hosts extremely difficult. With growing insight into these processes, however, the presented analysis may be refined to allow more precise predictions.

The amount of new proteins to be discovered by random expression cloning thus appears to be not as gigantic as originally thought, and from the total estimated enzymatic diversity of about 10^{13} distinct functional sequences (Burton et al., 2002) only a part is expected to be accessible in *E. coli*. Alternative hosts, particularly from taxonomic groups that contain only few expression signals functional in *E. coli* (e.g. *Streptomyces* sp. from the group of Actinobacteria), may yet broaden the range of exploitable genes.

4

Construction, characterization, and use of small-insert gene banks of DNA isolated from soil and enrichment cultures for the recovery of novel amidases

---

Esther M. Gabor, Erik J. de Vries, and Dick B. Janssen

Accepted for publication in *Environmental Microbiology*

To obtain new amidases of biocatalytic relevance, we used microorganisms indigenous to different types of soil and sediment as a source of DNA for the construction of environmental gene banks, following two different strategies. In one case, DNA was isolated from soil without preceding cultivation to preserve a high degree of (phylo)genetic diversity. Alternatively, DNA samples were obtained from enrichment cultures, which is thought to reduce the number of clones required to find a target enzyme. To selectively sustain the growth of organisms exhibiting amidase activity, cultures were supplied with a single amide or a mixture of different aromatic and non-aromatic acetamide and glycine amide derivatives as the only nitrogen source. Metagenomic DNA was cloned into a high-copy plasmid vector and transferred to *E. coli*, and the resulting gene banks were searched for positives by growth selection. In this way, we isolated a number of recombinant *E. coli* strains with a stable phenotype, each expressing an amidase with a distinct substrate profile. One of these clones was found to produce a new and highly active penicillin amidase, a promising biocatalyst that may allow higher yields in the enzymatic synthesis of β-lactam antibiotics.

1. INTRODUCTION

A major portion of microorganisms present in nature cannot readily be obtained in pure culture, which is why culture-independent techniques become more and more important for assessing diversity and exploiting the biocatalytic potential of microbial communities. The construction and screening of gene libraries prepared from DNA directly isolated from environmental samples is a recent and powerful tool for the discovery of new enzymes of biotechnological interest [e.g. lipases and esterases (Henne et al., 2000), chitinases (Cottrell et al., 1999), cellulases (Healy et al., 1995), polyketide synthases (Courtois et al., 2003), and amylases (Rondon et al., 2000); for a review see Lorenz et al. (2002)]. While standard methods based on the screening of isolated microorganisms are inherently limited to the tiny fraction of cultivable microbial species (<1 %; Torsvik et al., 2002), environmental gene banks in principle provide access to the entire sequence space present in nature (Handelsman et al., 1998).

One of the challenges of environmental cloning is the immense number of transformants that needs to be produced and screened. It has been estimated that more than 10^7 plasmid clones (5 kb inserts) or 10^6 BAC clones (100 kb inserts) would be required in order to represent the collective genomes of the thousands of different species that are typically present in a soil sample, assuming the idealized case of all species being equally abundant (Lorenz and Schleper, 2002). As a prerequisite to reach these large numbers, efficient and non-biased methods need to be employed for the extraction and purification of high-quality DNA from environmental samples (*Chapter 2*). The subsequent construction of gene banks from environmental DNA involves the same steps and techniques as the cloning of genomes of single microorganisms, i.e. fragmentation of genomic DNA, ligation into an appropriate vector system, and

transformation to a host organism, usually *E. coli* (Daniel, 2002). However, contaminants such as humic acids in the starting material, which are always co-extracted with nucleic acids from environmental samples like soil, affect enzymatic steps in the procedure (digestion of DNA and ligation) and decrease transformation efficiencies (Tebbe and Vahjen, 1993). Regarding the large numbers of clones that need to be produced, extreme care must therefore be taken to optimize each step in the used protocol.

In an ideal gene library with equal representation of all indigenous species and no non-productive clones, the number of clones that statistically need to be screened to find a positive one is solely determined by the frequency of organisms carrying one or several genes of interest in the source DNA. This frequency can be raised by a classical enrichment step preceding DNA isolation, where organisms are cultivated under selective pressure that favors the growth of bacteria expressing a desired activity. Like traditional strategies for enzyme discovery, this method may suffer from the fact that many organisms will not grow under laboratory conditions due to their special requirements in (nutritional) growth conditions. Many so-called "non-cultivable" bacteria, however, cannot be obtained as pure isolates, since they exist in nature as symbionts or as part of consortia, requiring the presence of certain other organisms for growth (Schink, 2002). Those organisms may be maintained in mixed liquid culture, making enrichment cultures a potential source of new genes from organisms that are not accessible by traditional techniques. Besides the anticipated decrease in gene bank size, the use of enrichment cultures as a starting material for cloning also facilitates DNA recovery and handling, thereby allowing faster access to the environmental gene pool. So far, this strategy has yielded new biotin biosynthesis operons (Entcheva et al., 2000) and genes conferring alcohol oxidoreductase activity (Knietsch et al., 2003b).

In this paper, we present our results from both direct environmental cloning and cloning from enrichment cultures for the isolation of new amidases, using growth selection of active clones on solid agar media. With these approaches, six clones comprising recombinant DNA with no or only low homology to existing database-entries were isolated. One of the positives expressed a novel penicillin amidase (penicillin acylase, EC 3.5.1.11) with highly favorable enzymatic properties, which demonstrates the potential of the applied approach to yield industrially relevant biocatalysts.

2. MATERIALS AND METHODS

2.1 Environmental samples

Inocula for enrichment cultures were obtained from 4 different locations in The Netherlands. Marine sediment was collected from the surface of daily-inundated saline mud flats during low tide near Paesens-Moddergat. Soil samples were taken from the upper 5 to 15 cm of the shore of a small goose pond, the lakeshore of the Lauwersmeer (sandy soil), and from a local agricultural field (loamy soil). The latter soil also served for DNA extraction without preceding cultivation.

2.2 Media and cultivation conditions

All selective media were based on minimal medium of pH 7.0. It contained per liter 5.3 g $Na_2HPO_4 \cdot 12H_2O$, 1.4 g KH_2PO_4, 0.2 g $MgSO_4 \cdot 7H_2O$, 1.0 g $(NH_4)_2SO_4$, 1 ml vitamin solution (Janssen et al., 1984), and 5 ml of a trace element solution. The trace element solution consisted of 780 mg $Ca(NO_3)_2 \cdot 4H_2O$, 200 mg $FeSO_4 \cdot 7H_2O$, 10 mg $ZnSO_4 \cdot 7H_2O$, 10 mg H_3BO_3, 10 mg of $CoCl_2 \cdot 6H_2O$, 10 mg $CuSO_4 \cdot 5H_2O$, 4 mg $MnSO_4 \cdot H_2O$, 3 mg $Na_2MoO_4 \cdot 2H_2O$, 2 mg $NiCl_2 \cdot 6H_2O$, and 2 mg $Na_2WO_4 \cdot 2H_2O$ per liter. In nitrogen-free minimal medium, $(NH_4)_2SO_4$ was replaced by Na_2SO_4, and $Ca(NO_3)_2 \cdot 4H_2O$ was omitted from the trace element solution. Solutions of carbon, nitrogen, and leucine sources were prepared in 50 mM potassium phosphate buffer (pH 7.0). All media used for gene bank screening and cultivation of isolated clones were supplied with 10 mg l^{-1} kanamycin. When solid LB-containing media were prepared, 15 g l^{-1} agar (Difco) was added. For solid media based on minimal medium, 15 g l^{-1} MP agarose (Roche) was used. Cells were grown at 30°C with shaking of liquid cultures at 200 rpm.

2.3 Enrichment cultures

Enrichments were carried out under nitrogen-limiting conditions, supplying one or several aromatic and aliphatic amides as a sole source of nitrogen. Inocula were prepared by suspending 5 g of environmental material in 50 ml enrichment medium (nitrogen-free minimal medium supplied with 10 mM glucose, 2.5 mM glycerol, and 2.5 mM sodium acetate as carbon sources). The mixtures were homogenized in a standard blender (Moulinex) for three 1-min intervals with 1 min breaks in-between to allow cooling, and incubated overnight at 30°C and 200 rpm for depletion of residual nitrogen. After settling of coarse material, sub-samples of 200 µl were used to inoculate 20 ml of enrichment medium, supplied with 1 mM D-phenylglycine amide as sole source of nitrogen for enrichment cultures ENR-M, ENR-S, and ENR-L. To ENR-G, an equimolar mixture of 10 different amides (Fig. 1.B) was added, providing each compound at a final concentration of 0.1 mM. After 2 days of incubation, 0.5 ml of cells were transferred to 50 ml of the same but fresh medium, and incubated for another 2 days before they were pelleted and used for DNA extraction.

2.4 DNA extraction

The protocol supplied with the Wizard Genomic DNA Purification Kit for DNA isolation from pure bacterial cultures (Promega) was adapted for the use of enrichment cultures. Cell pellets from enrichment cultures were resuspended in a total volume of 2.5 ml lysis buffer (50 mM Tris-HCl, 50 mM EDTA, 2 mg ml^{-1} lysozyme, and 0.1 mg ml^{-1} proteinase K; pH 7.0) and incubated at 37°C for 30 min. To the suspensions, 3 ml of Nuclei Lysis Solution (Promega) was added. After incubation at 80°C for 15 min, the mixtures were incubated at 65°C for 1 to 2 hours until complete lysis was reached. RNA was subsequently removed by adding 10 µl of DNAse-free RNAse solution (10 mg ml^{-1}) and incubating at 37°C for 30 min. Proteins were precipitated, using 1 ml of Protein Precipitation Solution (Promega), vortexing for 20 s at maximum speed and incubation on ice for 15 min, followed by a 10-min centrifugation step at 3,300 g. The supernatants were transferred to 0.1 volumes of sodium acetate (3 M, pH 5.0) and 2

volumes of absolute ethanol. Threads of precipitated genomic DNA were collected with a hooked Pasteur pipette and washed with 70 % ethanol. Nucleic acids were resuspended in 0.5 to 1 ml of 10 mM Tris buffer (pH 8.0) and used directly for gene bank construction.

For DNA isolation directly from soil, a cell extraction-based protocol, employing blending and differential centrifugation to remove cells from the sample matrix prior to lysis, was used as described previously (*Chapter 2*). DNA prepared this way needed to be further purified on 3 % (w/v) agarose gel (Roche), using the QIAEX II Gel Extraction System (Qiagen).

2.5 PCR-DGGE analysis

PCR-DGGE analysis was performed as described before (*Chapter 2*). Primers U968 (5'- AAC GCG AAG AAC CTT AC-3') and L1401 (5'-CGG TGT GTA CAA GAC CC-3') served to amplify a 402-bp section of bacterial 16S rRNA genes, including the highly variable V6 region (Engelen et al., 1995). Separation of PCR products by denaturing gradient gel electrophoresis (DGGE) was carried out with a device manufactured by Ingeny International BV (The Netherlands). Gels were silver stained (Bassam et al., 1991) and digitally photographed.

2.6 Gene bank construction

DNA fragments of 4 to 6 kb were prepared by mechanical shearing of DNA with a nebulizer (Invitrogen, catalog no. 7025-05) according to the instructions of the manufacturer. After ethanol precipitation in presence of 50 μg ml^{-1} glycogen as a carrier, DNA was resuspended and blunted with Klenow and T4 polymerase (Invitrogen). Fragments were purified from agarose gel using the QIAquick Gel Extraction Kit (Qiagen), and ligated into the *Eco*RV site of the high-copy plasmid vector pZero-2 (Invitrogen), using a 1:10 vector:insert ratio and a total amount of about 1 μg of DNA in a volume of 50 μl. T4 ligase (Invitrogen) was added to a final concentration of 0.1 U μl^{-1} and ligation mixtures were incubated at 16°C for 1 h before ethanol precipitation. Precipitated DNA was washed with 70 % ethanol and resuspended in 10 μl demineralized water. Aliquots of 2.5 to 5 μl were transformed to 80 μl of *E. coli* TOP10 cells [Δ(ara-leu)7697] by electroporation, using a 2-mm gap electroporation cuvette (Invitrogen) and a Gene Pulser apparatus (BioRad). Electrocompetent cells were prepared according to standard procedures (Sambrook et al., 1989) and allowed the production of $>10^9$ clones μg^{-1} supercoiled control plasmid (pSK^-) and 8,000 (loamy soil DNA) to 25,000 clones (enrichment DNA) when transforming ligation mixtures as described above.

Transformants were spread on LB agar plates containing 10 mg l^{-1} kanamycin for the selection of transformants carrying a pZero-2 derived recombinant plasmid. After incubation at 37°C for 1 day, colony forming units (cfu) were enumerated and at least 20 transformants were separately grown overnight in 5 ml LB medium. Plasmid DNA was isolated from these cultures using the High Pure Plasmid Isolation kit (Roche) and checked for insert size by enzymatic restriction with *Bam*HI and *Xho*I (Roche).

Average insert sizes were found to be 5.2 kb in all gene banks, with background levels of self-ligated vector molecules of ≤ 5 %.

2.7 Selection of amidase expressing clones

Transformants were washed from plates with minimal medium. Aliquots of the amplified gene banks were spread on minimal medium plates supplied with 0.2 % (w/v) glucose and 10 mg l^{-1} phenylacetyl-L-leucine or D-phenylglycine-L-leucine as the sole source of leucine. Colonies of amidase-expressing clones were visible after 3 days of incubation at 30°C. Aliquots of the gene bank derived from ENR-G were plated on 10 different nitrogen-free minimal media, containing 0.2 % (w/v) glucose, 5 mg l^{-1} leucine, and one of the amides used for enrichment (Fig. 1) as the only nitrogen source (250 µM final concentration). Positives could be scored after 3 to 5 days of growth at 30°C. Due to the use of amplified gene banks, active clones obtained after selection needed to be analyzed by enzymatic restriction to determine the number of unique transformants exhibiting the respective enzymatic activity.

2.8 HPLC analysis

All high-performance liquid chromatography (HPLC) analyses were carried out using a 10-cm Chrompack C18 column (0.5 cm diameter) in connection with Jasco PU-980 pumps and a Jasco MD-910 detector set at 214 nm. Compounds were isocratically eluted at a flow rate of 1 ml min^{-1} with a solution of 340 mg l^{-1} sodium dodecylsulfate and 680 mg l^{-1} $KH_2PO_4 \cdot 3H_2O$ in a 30:70 (v/v) acetonitrile/water mixture of pH 3.0 (adjusted with diluted phosphoric acid). Before injection, samples were quenched by 1:50 dilution in eluent.

2.9 Substrate profiling

All substrate conversions were carried out with cell-free extracts that were prepared from overnight cultures grown at 30°C and 200 rpm rotary shaking. Cells were harvested by centrifugation at 5,000 g for 10 min. Pellets were resuspended in 1/20 volume of ice-cold 50 mM potassium phosphate buffer (pH 7.0) containing 0.1 mM 1,4-dithiothreitol, and disrupted by sonification. Supernatants obtained after centrifugation at 5,000 g for 30 min at 4°C were used for the measurements. Protein concentrations of the cell-free extracts were determined by the method of Bradford. Enzymatic hydrolysis of the colorimetric substrates (Fig. 1.B, compounds 15-17) was followed by measuring the increase in absorbance at 405 nm caused by the release of 5-amino-2-nitrobenzoic acid ($\Delta\varepsilon_{405nm}$ = 9.09 mM^{-1} cm^{-1}) or *p*-nitroaniline ($\Delta\varepsilon_{405nm}$ = 13 mM^{-1} cm^{-1}), respectively, using a Perkin Elmer Lambda Bio 40 spectrometer. Conversion of non-chromogenic substrates was monitored by withdrawing samples of the reaction mixtures at different time points and analyzing them by HPLC. All substrates were supplied at a final concentration of 10 mM and reactions were carried out at 30°C in 50 mM potassium phosphate buffer of pH 7.0. Background activities caused by host strain proteins were determined with cell-free extract of *E. coli* TOP10 and subtracted from the activities measured with the recombinant *E. coli* strains.

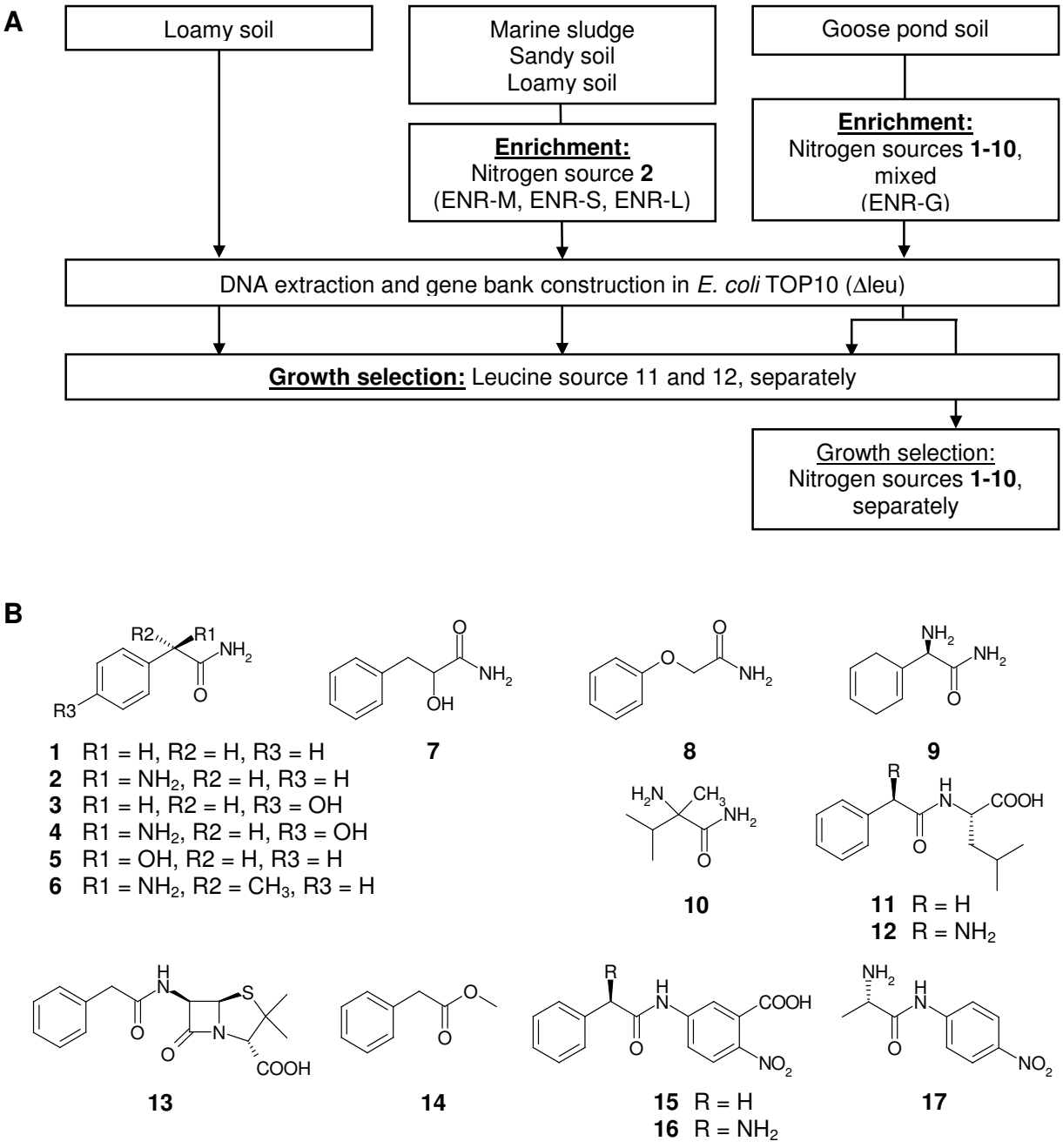


Figure 1. **A.** Strategy for the isolation of amidase-expressing clones. **B.** Substrates used for enrichments (1–10), selection media (1-12), and substrate profiling (2, 3, 5, 8, 9, 11, and 13-17). Phenylacetamide (1), D-phenylglycine amide (2), *p*-hydroxyphenylacetamide (3), *p*-hydroxy-D-phenylglycine amide (4), D-mandelamide (5), α-methylphenylglycine amide (6, racemic mixture), α-hydroxy-3-phenyl-propionamide (7, racemic mixture), phenoxyacetamide (8), D-dihydrophenylglycine amide (9), α-methylvaline (10, racemic mixture), phenylacetyl-L-leucine (11), D-phenylglycine-L-leucine (12), penicillin G (13), phenylacetic acid methyl ester (14), 2-nitro-5-[(phenylacetyl)amino]-benzoic acid (NIPAB, 15), D-2-nitro-5-[(phenylglycyl)amino]-benzoic acid (NIPGB, 16), alanine-*p*-nitroanilide (APNA, 17). While the conversion of compounds 1-14 was followed by HPLC, hydrolysis of the colorimetric compounds (15-17) could be observed photometrically at 405 nm.

2.10 Coding sequence determination and analysis

DNA insert sequences were determined at the Department of Medical Biology of the University of Groningen (The Netherlands) or BaseClear Holding BV (Leiden, The Netherlands) with at least two times coverage of each base. Open reading frames (ORFs) were identified with the "ORF Finder" service available at the NCBI web page (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). For translation to protein sequences, the Bacterial Code was selected, allowing ATG, GTG, TTG, ATT, and CTG as alternative start codons. The minimal ORF length was set to 270 bp. ORFs were translated and used as queries in blastp searches of the DDBJ/EMBL/GenBank database entries. When alternative start sites for the same gene were encountered, the largest ORF was selected. Only hits of amino acid stretches covering at least 20 % of the total protein sequence were recorded. Translated ORFs were analyzed for the presence of N-terminal signal sequences, using the SignalP program (Henrik et al., 1997).

The sequence data have been submitted to the DDBJ/EMBL/GenBank databases under accession numbers AY573296 (plasmid pM1), AY573297 (plasmid pS1), AY573298 (plasmid pS2), AY573299 (plasmid pG1), AY573300 (plasmid pL1), and AY573301 (plasmid pL2).

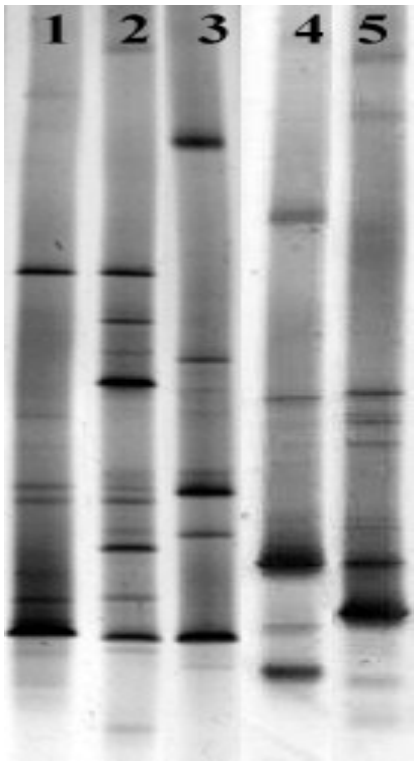


Figure 2. DGGE profiles of DNA isolated from enrichment cultures. Lane 1, ENR-M; lane 2, ENR-S; lane 3, ENR-L; lane 4, control culture for ENR-G (1 mM NH_4Cl added as sole source of nitrogen); lane 5, ENR-G.

3. RESULTS AND DISCUSSION

3.1 Diversity of source DNA

In this study, we were interested in the isolation of new enzymes that can convert derivatives of D-phenylglycine amide. These molecules are important intermediates in the biocatalytic synthesis of semi-synthetic β-lactam antibiotics, of which ampicillin, amoxicillin, cephalexin, and cefadroxil are prominent examples (Bruggink et al., 1998). To increase the chance of finding useful enzymes, we enriched marine sludge, sand, and loamy soil organisms on D-phenylglycine amide as sole nitrogen source, selecting organisms that can hydrolyze this compound and release ammonia (enrichments ENR-M, ENR-S, and ENR-L; Fig. 1.A). Glucose, glycerol, and sodium acetate were supplied as easily degradable carbon sources. To prevent the cultivation of only the fastest growing species and preserve a high degree of diversity of target organisms, the cultivation time was kept short in all experiments with only one transfer to fresh medium. As a measure to sustain diversity also in substrate specificity, a mixture of 10 different amides (Fig. 1.B, compounds 1-10) was used for the cultivation of goose pond microorganisms (ENR-G). All enrichments were carried out under nitrogen-limiting conditions, which was confirmed by the absence of growth in parallel cultures to which no nitrogen source was added. Consequently, the selection of nitrogen-fixing organisms under the applied conditions can also be excluded.

To estimate the bacterial diversity in the different cultures, 16S rRNA fragments were PCR-amplified from the isolated DNA and separated by denaturing gradient gel electrophoresis (DGGE) (Fig. 2). Depending on the inocula, 10 to 16 different bands were observed, indicating the presence of relatively low numbers of abundant organisms. In contrast, a much higher phylogenetic diversity (44 bands) was detected when analyzing metagenomic DNA directly extracted from the same loam soil sample used as the inoculum for enrichment ENR-L (*Chapter 2*). This shows that a major portion of bacterial species was lost during enrichment, which may be due to the inability of certain species to metabolize the provided nitrogen source (intended effect) or due to the applied general growth conditions (adverse effect), e.g. cultivation in liquid medium, supplied growth substrates, temperature, aeration, or competition between numerically inferior and superior bacterial populations (Santegoeds et al., 1996). To distinguish between these two situations, PCR-DGGE analysis was also carried out for a control culture that was set up in the same way as ENR-G, except that no selective nitrogen source but readily accessible NH_4Cl was added. Fig. 2 (lanes 4 and 5) reveals a similar degree of diversity with clear differences concerning the location and intensity of bands in the two DGGE patterns, which suggests that most of the enriched organisms were able to grow because of their capacity to (co-)metabolize the selective compounds. This conclusion is supported by an almost complete (>90 %) depletion of amides 1-4 and 7-9 (Fig. 1.B) in ENR-G, as determined by HPLC analysis of the culture medium before inoculation and after cell collection. The chemical stability of the supplied amides under the assay conditions was such that no spontaneous hydrolysis occurred.

The degradation of D-mandelamide, α-methylphenylglycine amide, and α-methylvaline could not be monitored due to overlapping substrate and product peaks. Complete degradation of the supplied nitrogen source phenylglycine amide was also observed for ENR-M, ENR-S, and ENR-L.

3.2 Gene bank construction

For DNA extraction from enrichment cultures, we used an optimized protocol that allowed the parallel processing of all four samples and yielded large amounts of high molecular weight DNA with no need for further purification. Even organisms that are known to be difficult to lyse like *Rhodococcus* or *Mycobacterium* sp. were readily accessible as established in preliminary experiments. Enrichment DNA and nucleic acids isolated from loamy soil without preceding cultivation was used to construct five different gene banks (Table 1).

Table 1. Diversity of source DNA and gene bank size

DNA source[a]	No. of DGGE bands, *n*	Minimal gene bank size[b], N_{min}	Actual gene bank size	No. of positives (plasmid names)
ENR-M	11	18,000 (31,000)	23,000	1 (pM1)
ENR-S	14	23,000 (40,000)	30,000	2 (pS1 and pS2)
ENR-L	10	17,000 (29,000)	35,000	0
ENR-G	16	27,000 (46,000)	25,000	2 (pG1 and pG2)
Loamy soil	44	60,000 (126,000)	80,000	2 (pL1 and pL2)

[a] As inocula for enrichments, marine sludge (ENR-M), sand soil (ENR-S), loamy soil from an agricultural field (ENR-L), and soil collected at a goose pond (ENR-G) were used. From the loamy soil, DNA was also extracted without preceding cultivation (last row).

[b] The number of minimally required clones was calculated by Eq. 1 with $P = 0.9$, $I = 5.2$ kb, $X = 0.9$ kb, and $G = 3{,}100$ kb, assuming gene expression from native promoters and even distribution of the different genomes. The numbers in brackets were calculated with $X = 2.7$ kb, which is the typical gene size of penicillin amidases.

In a first step, high molecular weight DNA was mechanically fragmented using a nebulizer (Bodenteich et al., 1994). In this small plastic device, DNA solutions are squeezed through small pores by overpressure, creating shearing forces that cause DNA strands to break. By adjusting pressure and shearing time, we were able to readily produce fragments between 1 and 10 kb at will with size distributions that were more narrow than those obtained in enzymatic digests (data not shown), which allows the efficient use of a possibly limited amount of source DNA. A further advantage of mechanical compared to enzymatic fragmentation is the fact that it is truly random and not dependent on restriction sites or methylation patterns, which may bias genome representation in the produced gene bank (Oefner et al., 1996). Furthermore, enzymatic restriction may be inhibited by contaminants in the DNA extract, which can especially be a problem when working with nucleic acids directly isolated from environmental samples. Fragments were size-selected on gel, blunted with Klenow and T4

polymerase, and ligated into the *Eco*RV site of plasmid pZero-2, a high-copy number vector. A lethal gene (*ccdB*) that is disrupted and, consequently, not expressed when an insert is incorporated, allowed the construction of gene banks with background levels of self-ligated vector of ≤5 %. Contaminants present in the DNA extract prepared from loamy soil could not be entirely removed during gel purification, which only allowed the production of about 30 % of the number of clones that was produced from a corresponding amount of the more pure enrichment DNA. In average, gene banks contained inserts of 5.2 kb.

3.3 Estimation of the minimally required gene bank size

The number of clones that needs to be produced to cover the collective genomes present in a DNA extract can be estimated by means of DGGE patterns. This is particularly true when working with DNA isolated from an enrichment culture that is of solely prokaryotic origin. DNA extracts obtained from various environmental samples by direct lysis methods, in contrast, have been shown to contain major portions (> 60 %) of eukaryotic DNA (*Chapter 2*), leading to an increase in the required library size due to the formation of non-productive clones with eukaryotic inserts, virtually diluting the desired prokaryotic recombinants. Although this number can often be limited by the use of cell extraction-based DNA isolation protocols, a certain fraction of eukaryotic DNA is always co-extracted.

Based on the assumption that expression of environmental genes is mostly dependent on native promoters (Daniel, 2002) and that bacterial species are equally abundant in the DNA extract, the minimal gene bank size N_{min} can be calculated by:

$$N_{min} = \frac{\ln(1-P)}{\ln(1-\frac{I-X}{G \cdot n})} \qquad \text{Equation 1}$$

Here, P is the probability to find each environmental gene at least once, n is the number of different species present, I is the average insert size in the gene bank, X is the average gene size, and G is the average genome size. Whereas X can be neglected when the average insert size is much bigger (e.g. in BAC libraries), it needs to be taken into account for small insert gene banks as constructed in this study. In view of most microorganisms not being discovered yet, parameters X and G can only be roughly estimated on basis of already sequenced and annotated organisms. When analyzing the chromosome sequences of all 95 completely sequenced prokaryotic organisms currently available at the TIGR database, neglecting alternative species of the same genus, G was found to be 3,100 kb. The average length of genes in the database (X) is 0.9 kb. While P can be chosen and I is adjusted in the procedure at will, n needs to be determined experimentally.

In a DGGE profile, distinct bands correspond to 16S rRNA fragments of different melting behavior resulting from sequence differences, which are in turn caused by a different phylogenetic origin. Consequently, individual PCR fragments represent unique bacterial populations (and possibly different 16S rRNA genes within a single species) and can be used as a rough estimate for n. It needs to be taken into account

that PCR-DGGE only reveals populations that account for more than 0.1 % of the total community (Gelsomino et al., 1999), thereby failing to detect the majority of bacterial species present in e.g. soil samples. However, as organisms of lower abundance are unlikely to be significantly represented in small insert gene banks as typically constructed nowadays ($\leq 10^6$ clones), DGGE profiles still provide useful information on the number of different genomes that will constitute the major part of a resulting gene bank.

As shown in Table 1, all enrichment gene banks were (nearly) complete with respect to the calculated minimal gene bank sizes and allowed the isolation of 1 or 2 clones exhibiting the desired activity. In principle, one would expect to recover at least one positive clone per species that passed the selection procedure (i.e. 10 to 16), provided that a suitable screening method is used. However, expression of the active protein may fail due to the lack of proper expression signals, incorrect folding or processing, non-availability of crucial cofactors, etc. Furthermore, some of the detected organisms may have only been able to withstand the selective pressure by using nitrogen-containing compounds or excess ammonia released by the primary degraders of the provided amides, not containing any active amidase themselves. Another reason for the apparent discrepancy, however, can be found in the calculation of N_{min} itself, which is based on the idealized assumption that all species are equally abundant in the source DNA. As can be seen from the DGGE profiles, band intensities though vary by more than a factor 10, meaning that N_{min} considerably underestimates the number of clones that is required to represent the collective genomes. Another parameter that strongly influences the outcome of Eq. 1 is X. While the average gene size in microorganisms is relatively small, specific classes of coding sequences can be much larger. Penicillin acylases, our primary targets, for instance, are typically 3 times larger, leading to minimal library sizes that are significantly bigger due to the decreased chance to clone a complete gene on DNA inserts of a given size (Table 1, in brackets). Therefore, more recombinants would need to be screened for a comprehensive mining of the biocatalytic potential of organisms present in the enrichment cultures.

3.4 Recovery of clones with amidase activity

Clones exhibiting amidase activity were selected on a medium containing phenylacetyl-L-leucine or D-phenylglycine-L-leucine (Fig. 1.B, compounds 11 and 12) as the sole source of leucine, allowing only the growth of recombinants that can complement the leucine auxotrophy of the *E. coli* TOP10 host strain by hydrolyzing the selective amide (Forney and Wong, 1989). As those compounds resemble D-phenylglycine amide, the selective nitrogen source used for cultivation of marine sludge, sand, and loam soil organisms, we expected to recover most of the amidases that conferred the ability to grow to their host organisms. The loamy soil gene bank produced without cultivation was screened accordingly.

For enrichment culture ENR-G, a mixture of different N-sources was employed, which possibly sustained the growth of organisms carrying enzymes with substrate specificities other than for phenylacetylated or D-phenylglycylated compounds. Besides leucine-limited medium, we therefore used nitrogen-limited agar media, in

which amides 1-10 (Fig. 1) were separately supplied as the sole nitrogen source. In this way, two clones were found in gene bank ENR-G, one growing the best on D-phenylglycine amide-containing medium (plasmid pG1) and the other on D-mandelamide-containing medium (plasmid pG2). Activity of the latter clone, however, was lost when it was transferred to fresh medium, leading to its exclusion from further analysis. From the other gene banks, 4 unique clones with a stable amidase phenotype were isolated on phenylacetyl-L-leucine, and one (plasmid pS1) was found on the Cα-amino derivative (Table 1). To assure that only plasmid-encoded activity was selected, recombinant plasmids were isolated, retransformed to *E. coli* TOP10, and plated on the selective medium on which they were originally isolated.

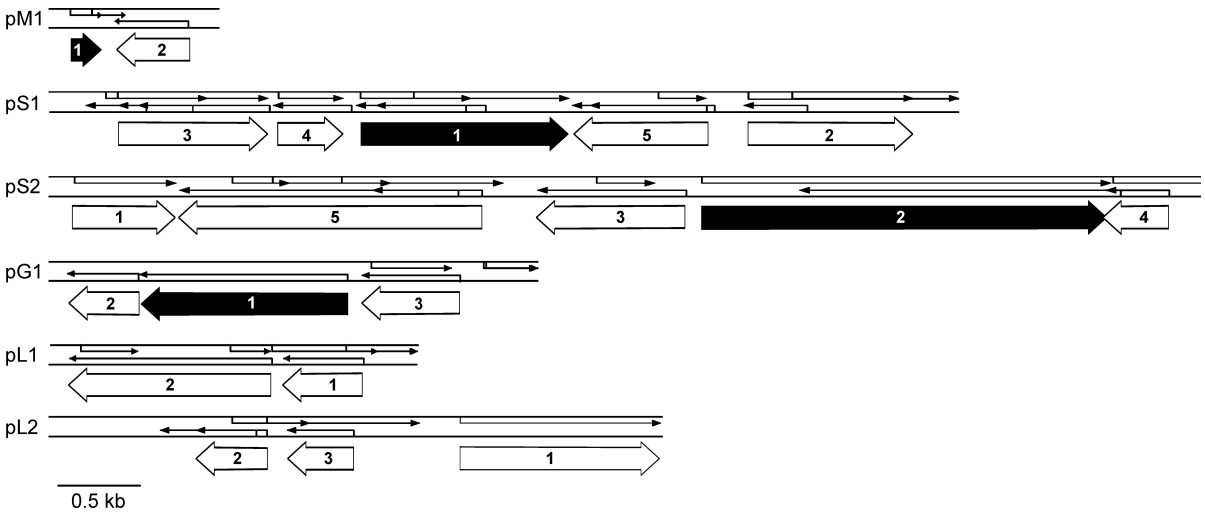


Figure 3. Genetic organization of environmental insert DNA encoding enzymes with amidase activity. The promoter of the vector (P_{lac}) is located on the left side (upper strand) of the shown inserts. Black arrows indicate ORFs that are responsible for the observed enzyme activity. ORF numbers refer to those given in Table 2.

3.5 Sequence analysis of the recombinants

The six amidase activity-exhibiting clones were shown to contain a total of 50 heterologous ORFs with the major part having no homology to any GenBank database entry (Fig. 3). Only 20 ORFs yielded database hits with moderate (25 %) to high (91 %) identity scores, including 7 predicted coding sequences of unknown function (Table 2). No single hit revealed complete identity with a known protein, which illustrates the potential of the approach taken for accessing unexplored sequence space. In 4 cases, the ORF responsible for the observed enzyme activity could easily be identified due to the found homologies, whereas pL1 and pL2 did not contain obvious candidates among their potential protein-coding regions. For the verification of the enzyme-coding sequences by subcloning, we therefore focused on pM1, pS1, pS2, and pG1. On pM1, a putative hydrolase homologue was found, preceded by the vector-encoded P_{lac} promoter and a good ribosome binding site, which made the respective ORF the most likely amidase-coding region although the resulting enzyme would be N-terminally truncated compared to its database relative. The activity of *E. coli* (pS1) on D-

phenyglycine-L-leucine, a molecule that is structurally very similar to a dipeptide, indicated the expression of the dipeptidase homologue encoded by ORF 1. Similarly, the substrate profiles of *E. coli* (pS2) and *E. coli* (pG1) supported the assumption that their assayed activities were caused by the encoded penicillin acylase and mandelamide hydrolase homologues, respectively (Table 3).

Table 2. Amidase activity conferring plasmids, identified ORFs, and sequence similarities

Plasmid (insert size, ORF no.[a])	ORF[b] (no. of encoded amino acids)	Closest database homologue (accession number)	Organism	Identity[c]
pM1 (1.017 kb, 3)	**ORF1** (92)	**putative hydrolase (NP_461274)**	*Salmonella typhimurium*	91 %, 67/199
	ORF2 (243)	putative transporter (NP_416795)	*Escherichia coli*	87 %, 233/610*
pS1 (5.461 kb, 17)	**ORF1** (443)	**probable Xaa-Pro dipeptidase (NP_419119)**	*Caulobacter crescentus*	41 %, 427/428*
	ORF2 (348)	TonB-dep. receptor (NP_420476)	*Caulobacter crescentus*	58 %, 331/809*
	ORF3 (318)	SalR transcriptional repressor (NP_639440)	*Xanthomonas campestris*	37 %, 282/361
	ORF4 (136)	hypothetical protein (B87530)	*Caulobacter crescentus*	65 %, 96/101*
	ORF5 (285)	hypothetical protein (SCE20_22)	*Streptomyces coelicolor*	46 %, 282/290
pS2 (7.019 kb, 12)	ORF1 (215)	hypothetical protein (ZP_00050555)	*Magnetospirillum magnetotacticum*	46 %, 178/287
	ORF2 (874)	**penicillin G amidase (AAP20806)**	*Achromobacter xylosoxidans*	84 %, 822/843*
	ORF3 (319)	transcr. activation protein (P52686)	*Pseudomonas* sp.	31 %, 282/306
	ORF4 (135)	hypothetical protein (ZP_00118494)	*Cytophaga hutchinsonii*	63 %, 110/612
	ORF5 (648)	ABC transporter (NP_437604.1)	*Sinorhizobium meliloti*	52 %, 602/604*
pG1 (2.933 kb, 6)	**ORF1** (495)	**mandelamide hydrolase (AAO23019)**	*Pseudomonas putida*	68 %, 466/507
	ORF2 (168)	benzaldehyde dehydrogenase (AAO23020)	*Pseudomonas putida*	79 %, 143/436*
	ORF3 (233)	putative regulator (AAO23018)	*Pseudomonas putida*	63 %, 149/233
pL1 (2.216 kb, 6)	ORF1 (203)	probable peroxidase (G64843)	*Escherichia coli*	33 %, 190/243*
	ORF2 (512)	hypothetical protein (NP_661777)	*Chlorobium tepidum*	29 %, 454/522
pL2 (3.681 kb, 6)	ORF1 (451)	hypothetical protein (NP_772653)	*Bradyrhizobium japonicum*	25 %, 271/1195*
	ORF2 (152)	hypothetical protein (NP_442421)	*Synechocystis* sp.	37 %, 114/148
	ORF3 (141)	anthranilate phosphoribosyl-transferase (NP_718589)	*Shewanella oneidensis*	40 %, 76/347*

[a] ORFs with alternative start sites for the same gene were counted as one and are represented by the largest ORF.

[b] ORFs that are not listed did not show significant homology to any database entry. ORFs marked in bold were identified to be responsible for the observed enzyme activity.

[c] The numbers behind the comma indicate the length of the homologous amino acid sequence stretch with respect to the protein length of the database hit. Protein sequences derived of ORFs that are marked by a star (*) were predicted to carry a signal sequence, using SignalP software (Henrik et al., 1997).

Subclones carrying the above-mentioned ORFs in pZero-2 displayed the same activities as the original recombinants, identifying the selected sequences indeed as the enzyme-coding regions. Analysis of the deduced protein sequences suggested that the dipeptidase and the penicillin acylase homologue carry N-terminal signal sequences that may lead to their secretion into the periplasmatic or extracellular space (Table 2).

3.6 Substrate profiling

To explore the biocatalytic potential of the retrieved enzymes, their relative activities were determined on a number of different substrates using cell-free extracts (Table 3). All extracts showed no or only low activity towards compounds with an aliphatic acyl moiety, which is not surprising regarding the aromatic substrates used for selection of active clones. Enzymes encoded by plasmid pS1 and pL1 appeared to not accept substrates with a substituent on the Cα-position of the acyl moiety, although *E. coli* (pS1) could grow on D-phenylglycine-L-leucine. Also the pM1 encoded enzyme preferred substrates with no substituent at this position. Recombinants carrying pG1, pS2, and pL2, in contrast, could readily convert α-amino compounds with pG1 and pS2 even being able to accept a α-hydroxy group. While *E. coli* (pG1) was most active on D-phenylglycine amide, all other enzymes hydrolyzed the chemically more labile ester bond of phenylacetic acid methyl ester the fastest.

Table 3. Relative activities of the 6 recovered environmental clones on different substrates

	Relative activity of clones carrying a recombinant plasmid [%][b]					
Compound[a]	pM1	pS1	pS2	pG1	pL1	pL2
D-Phenylglycine amide (2)	10	-	56	100	-	38
p-Hydroxy-D-phenylglycine amide (3)	33	-	65	37	-	52
D-Mandelamide (5)	-	-	46	33	-	-
Phenoxyacetamide (8)	-	-	4	34	-	34
D-Dihydrophenylglycine amide (9)	-	-	-	7	-	-
Phenylacetyl-L-leucine (11)	82	100	79	7	20	45
Penicillin G (13)	15	22	83	-	-	-
Phenylacetic acid methyl ester (14)	100	100	89	41	100	100
NIPAB (15)	64	33	100	-	-	-
NIPGB (16)	27	-	46	-	-	-
APNA (17)	9	11	2	-	-	-

[a] Numbers in brackets refer to Fig. 1.B.

[b] Measurements were carried out with cell-free extracts of the respective clones, supplying substrates at a final concentration of 10 mM. Relative activities were calculated with respect to the compound that was converted the fastest by each extract (1.1 mU, pM1; 0.9 mU, pS1; 31.2 mU, pS2; 47.1 mU, pG1; 0.5 mU, pL1; 0.8 mU, pL2) with 1 mU defined as the conversion of 1 nmol substrate min^{-1} mg $protein^{-1}$. -, no conversion.

Concerning the nature of the leaving groups of their substrates, enzymes fell into two categories, one being solely active on primary amides or amides carrying a small aliphatic leaving group (pG1, pL1, and pL2), and another group also hydrolyzing substrates with more bulky leaving groups, i.e. penicillin G and the colorimetric substrates (pM1, pS1, and pS2). Of the latter group, only *E. coli* (pS2) was found to be capable of both hydrolyzing penicillin G and catalyzing a condensation reaction when incubated with 6-aminopenicillanic acid and phenylacetamide, yielding penicillin G (Fig. 4).

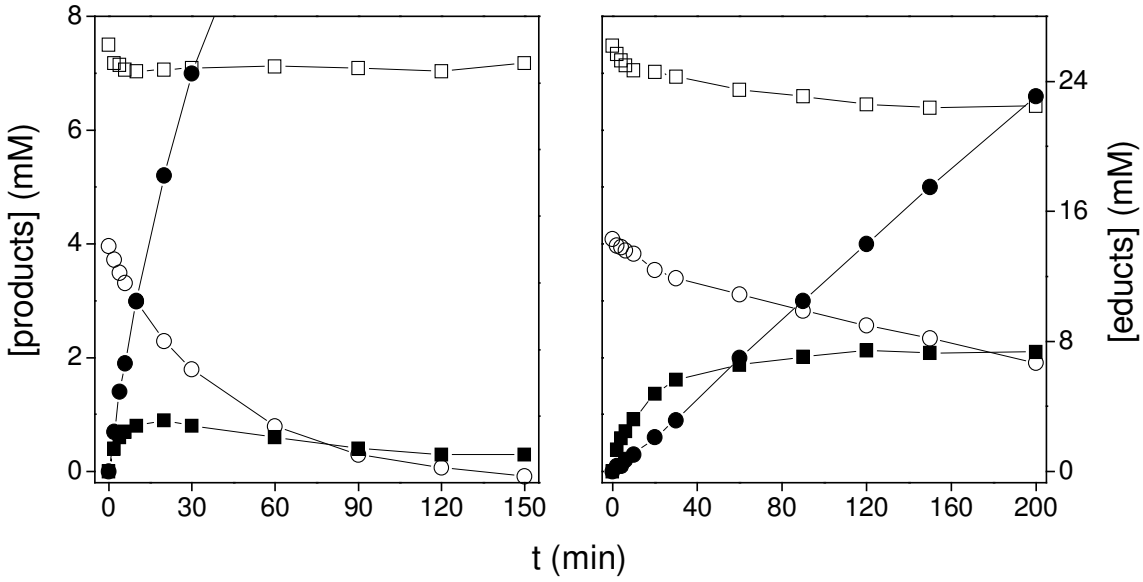


Figure 4. Biocatalytic synthesis of penicillin G by the penicillin amidase encoded on pS2 (right) and the one of *E. coli* ATCC 11105, carried by the plasmid pEC (left). Cell-free extracts of *E. coli* TOP10 (pS2) and *E. coli* HB101 (pEC) were prepared as described in Materials and Methods, except that *E. coli* HB101 cells were grown at 17°C for 3 d with 0.1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) added for induction. Extracts were incubated with 15 mM phenylacetamide and 25 mM 6-aminopenicillanic acid at 30°C. Educt and product concentrations in the course of the reaction were determined by HPLC analysis. The pEC plasmid was kindly provided by DSM (Delft, The Netherlands). Symbols: phenylacetamide (○), 6-aminopenicillanic acid (□), phenylacetic acid (●), penicillin G (■).

Two side reactions, however, were also found to take place when synthesizing penicillin G, namely the hydrolysis of phenylacetamide and the hydrolysis of the formed antibiotic, which is typical for penicillin amidase-catalyzed antibiotic synthesis reactions with activated side-chain donors (Kasche, 1986). As a consequence, the concentration of antibiotic passes through a maximum in time, of which the height is solely determined by the substrate concentrations and the kinetic properties of the enzyme (Youshko and Svedas, 2000).

Compared to the well-studied *E. coli* penicillin amidase, the pS2-encoded enzyme allowed a more than 2-fold higher maximal level of penicillin G accumulation (Fig. 4). The performance of the new biocatalyst was also significantly better in the production of the clinically relevant antibiotics ampicillin and amoxicillin, which makes the recovered enzyme a potential candidate for biotechnological application.

In conclusion, cloning from enrichment cultures was found to be an efficient tool for enzyme discovery although cultivation steps preceding DNA extraction may limit the access to certain parts of the microbial gene pool. Taking into account that virtually all experimental steps from DNA extraction to construction and transformation of recombinant plasmids can be performed more efficiently and with less effort, the preparation of several small enrichment gene banks of limited diversity may be preferable compared to the construction of a single exhaustive gene bank without cultivation, especially when time and resources are limited.

ACKNOWLEDGEMENTS

We thank M. van der Velde and S. den Dulk from Bioclear (Groningen, The Netherlands) for DGGE analysis.

5

A novel penicillin acylase from the environmental gene pool with improved synthetic properties

Esther M. Gabor, Erik J. de Vries, and Dick B. Janssen

Accepted for publication in *Enzyme and Microbial Technology*

A new penicillin acylase was isolated by cloning and functional screening of DNA isolated from a sand soil enrichment culture. Sequence analysis of this enzyme, PAS2, revealed homology to a group of prominent penicillin G acylases, including the intensively studied enzyme of *E. coli* ATCC 11105. Accordingly, PAS2 was found to be an Ntn-hydrolase with an N-terminal serine as the catalytic nucleophile, located on its 61.9 kDa β-subunit. The α-subunit was shown to have a molecular mass of 25.5 kDa. To evaluate the biocatalytic performance of the new enzyme, the complex kinetic parameters α, β_0, and γ were determined for the kinetically controlled synthesis of a number of important semi-synthetic penicillins and cephalosporins. While α is a measure for the relative affinity of the enzyme for the activated acyl donor, β_0 and γ quantify the efficiency of acyl-transfer to the β-lactam nucleophile. Compared to the penicillin acylase of *E. coli*, PAS2 showed superior potential for the synthesis of 6-aminopenicillanic acid-derived antibiotics, allowing the accumulation of up to 2.3-fold more target product at significantly higher conversion rates. In the synthesis of amoxicillin, for instance, 1.6-fold more antibiotic was formed using the new enzyme, making PAS2 an interesting candidate for biocatalytic application.

1. INTRODUCTION

Penicillin acylase (penicillin amidase, PA, EC 3.5.1.11) occurs in many bacteria, yeasts and filamentous fungi. Since more than a decade, this enzyme is commercially employed for the large-scale hydrolysis of penicillin G that is fermentatively produced by *Penicillium chrysogenum* (Bruggink et al., 1998; Demain, 2000; Shewale et al., 1990; Valle et al., 1991). The reaction product, 6-aminopenicillanic acid (6-APA), is a key intermediate in the synthesis of clinically useful semi-synthetic penicillins such as ampicillin or amoxicillin.

While the hydrolytic application of PA is thus well established, the synthetic capacity of the enzyme, although already described some forty years ago (Cole, 1969), is still less exploited due to thermodynamic constraints. Since the reaction equilibrium is unfavorable, yields in the direct condensation of 6-APA with D-phenylglycine (D-PG) and its derivatives are very low (Kasche, 1986). The same holds true for the production of semi-synthetic cephalosporins such as cephalexin and cefadroxil with 7-aminodesacetoxycephalosporanic acid (7-ADCA) as the β-lactam group.

Yields can be improved by using activated acyl donor moieties, mostly the amide or ester derivatives of the acid (Fig. 1). In these kinetically controlled reactions, significantly higher antibiotic concentrations can transiently be reached during the conversion process. However, yields are still limited due to two enzyme-catalyzed side-reactions: (1) the hydrolysis of the activated acyl donor and (2) the hydrolysis of the synthesized antibiotic. Due to the undesired hydrolytic reactions, the unproductive loss of acyl donor exceeds the accumulation of antibiotic in the course of the conversion, which is a major drawback for an industrial process. Many studies have been aimed at improving the kinetically controlled synthesis of semi-synthetic β-

lactam antibiotics by medium engineering or modifying the reaction conditions. This includes optimization of the pH (Youshko et al., 2002b), addition of cosolvents (Fernández-Lafuente et al., 1996), or the use of high substrate concentrations (Youshko et al., 2001) as well as improvement of the biocatalyst (e.g. by immobilization; Alvaro et al., 1990).

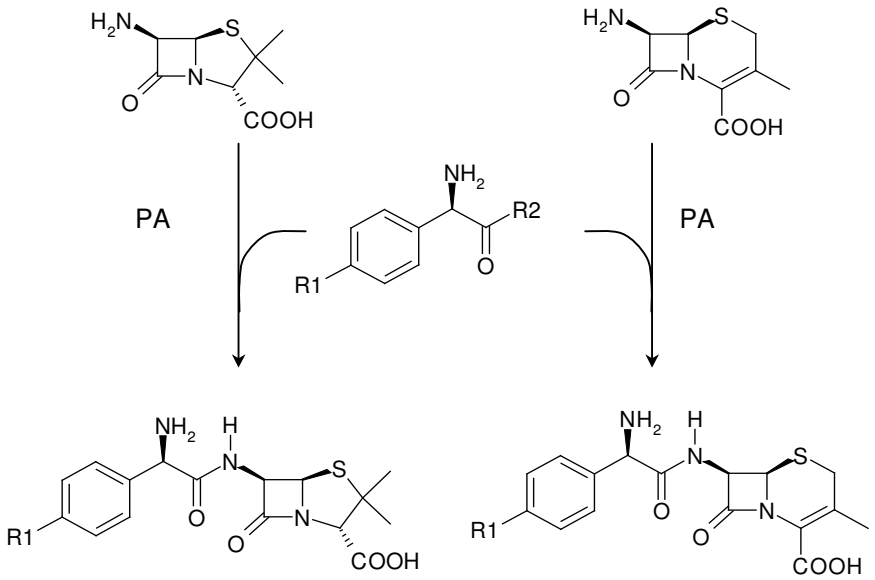


Figure 1. The use of penicillin acylase in the synthesis of semi-synthetic penicillins (left; ampicillin, R1 = H; amoxicillin, R1 = OH) and cephalosporins (right; cephalexin, R1 = H; cefadroxil, R1 = OH). In kinetically controlled synthesis reactions, an activated derivative of the acyl-donor is used, most commonly the amide (R2 = NH_2) or methyl ester (R2 = OCH_3) of phenylglycine (R1 = H) and *p*-hydroxyphenylglycine (R1 = OH), respectively.

A more fundamental approach to enhance product yields is the use of new biocatalysts with improved kinetic properties, since it has been shown that kinetically controlled reactions are fundamentally influenced by the kinetic parameters of the employed enzyme (Alkema et al., 2002b; Alkema et al., 2002c; Youshko et al., 2002a). The availability of a penicillin acylase with outstanding synthetic performance is thus a key factor in the development of biocatalytic processes for the synthesis of semi-synthetic β-lactam antibiotics that are competitive with traditional chemical condensation approaches.

Here, we describe the synthetic properties of a new penicillin acylase, PAS2, which was obtained by functional screening of an environmental gene bank constructed from a sand soil enrichment culture. Initial experiments revealed a higher preference towards the synthesis reaction for semi-synthetic penicillins as compared to the enzyme isolated from *E. coli* ATCC 11105, which is the best studied penicillin G acylase to date and which we use as a benchmark (Alkema et al., 2000, Arroyo et al., 2003). In this paper, we present a comprehensive kinetic study of this interesting new enzyme and show that its use in the production of semi-synthetic penicillins could lead to more efficient processes than with *E. coli* PA.

2. MATERIALS AND METHODS

2.1. Cloning of pas2 from a sandy soil enrichment culture

The construction of environmental gene banks and the screening procedure for amidase-producing transformants is in detail described in *Chapter 4*. Briefly, organisms present in a sandy soil sample were grown under selective pressure, supplying D-phenylglycine amide (PGA) as a sole source of nitrogen. Genomic DNA was isolated from this enrichment culture and used to construct an environmental gene bank in the leucine-auxotroph host strain *E. coli* TOP10, using plasmid pZero-2 (Invitrogen) as a vector. Selection of clones able to utilize D-phenylacetyl-L-leucine as a source of leucine (Forney and Wong, 1989) was carried out on selective agar plates and yielded two different recombinants. One of the clones also exhibited activity towards 2-nitro-5-[(phenylacetyl)amino]-benzoic acid (NIPAB), a well-known colorimetric substrate of PAs (Kutzbach and Rauenbusch, 1974). Sequencing and further substrate profiling revealed that S2 indeed encoded a PA, which was named PAS2. The DNA sequence of the *pas2* gene was submitted to GenBank as accession number AY573298.

2.2. Subcloning of the pas2 gene

To achieve high-level expression of PAS2, *pas2* was cloned behind the tightly regulatable P_{BAD} promoter located on pBAD/*Myc*-HisA_*Nde*I. This plasmid vector solely differs from the commercially available vector pBAD/*Myc*-HisA (Invitrogen) by carrying an *Nde*I instead of an *Nco*I recognition sequence in its multiple cloning site. The *pas2* gene was cloned including its own signal sequence, using a forward PCR primer based on the 5'-end of the gene with an introduced *Nde*I restriction site (underlined, start codon bold), 5' ttggagacagag<u>cat</u>**atg**aagcagcatttgttg 3', and a reverse primer based on the 3'-end of the gene with a *Sal*I site (underlined) incorporated (5' ccagggc<u>gtcgac</u>acggtcagtagcg 3'). PCR amplification was carried out with pWO polymerase (Roche) under standard conditions, using whole cells of the original clone S2 as a template. PCR product and vector were digested with *Nde*I/*Sal*I and *Nde*I/*Xho*I, respectively, and ligated with T4 ligase according to the instructions of the manufacturer (Roche). The ligation mixture was transformed to electrocompetent *E. coli* TOP10 cells, and the construct (pBADPAS2) was confirmed by sequencing.

2.3. Protein purification

E. coli PA was purified as described before (Alkema et al., 2000). The obtained enzyme solution was concentrated by ultrafiltration (Amicon bioseparations, YM 30 filter) before the enzyme was rebuffered in 50 mM potassium phosphate buffer (pH 7.0) with 5 % glycerol, using an Econo-pac 10DG column (BioRad).

For PAS2, a similar purification scheme was used. *E. coli* TOP10 (pBADPAS2) cells were grown in LB (Sambrook et al., 1989) at 17°C with rotary shaking at 200 rpm. To induce protein expression from P_{BAD}, the medium was supplied with 0.8 % arabinose after 2 days of growth. After another 24 h of incubation, cells were harvested by centrifugation at 5,000 g for 10 min (4°C). To prepare a periplasmatic extract, cells were resuspended in 1/10 of the original culture volume of ice-cold osmotic shock

buffer (20 % sucrose, 100 mM Tris-HCl, 10 mM EDTA; pH 8.0) and centrifuged as described above. Cell walls were disrupted by resuspending the cell pellet in 1/10 of the original culture volume of ice-cold 1 mM EDTA. After centrifugation (6,000 g, 15 min, 4°C), 1 M potassium phosphate buffer (pH 7.0) was added to the supernatant (periplasmatic extract) to a final concentration of 50 mM. Subsequently, solid $(NH_4)_2SO_4$ was used to adjust a final concentration of 1.5 M, while stirring at 4°C. The solution was subjected to hydrophobic interaction chromatography, using a Resource Phe column (Amersham Pharmacia Biotech), and eluted with a linear gradient of 1.5 M to 0 M $(NH_4)_2SO_4$ in 20 mM potassium phosphate buffer (pH 7.0). PAS2 eluted at a concentration of 300 mM $(NH_4)_2SO_4$. Fractions containing enzyme activity were pooled, concentrated and rebuffered as described for *E. coli* PA. About 10 mg PAS2 could be obtained per liter of culture grown as explained above. The purity of the enzyme was > 95 % as judged by SDS-PAGE. The enzyme was stored at –20°C and could be defrosted several times without detectable loss of activity.

The amount of active enzyme in penicillin acylase preparations was determined by titration with the irreversible inhibitor phenylmethylsulfonyl fluoride (Roche) as done by Alkema et al. (1999).

2.4. Mass spectrometry

The molecular masses of the two subunits of PAS2 were determined at the Mass Spectrometry Core Facility, University of Groningen (The Netherlands), using an electrospray triple quadrupole mass spectrometer (API 3000, PE-Sciex). Full-scan spectra were recorded with a step size of 0.1 amu and analyzed with Biomultiview software (version 1.5, PE-Sciex). For this experiment, the buffer system of the original enzyme solution was replaced by a 10 mM ammonium acetate buffer (pH 6.8) with an Econo-pac 10DG column (BioRad) and 0.1 % formic acid was added before analysis.

2.5. Kinetic measurements

All enzymatic conversions were carried out in 50 mM potassium phosphate buffer (pH 7.0) at 30°C. Steady-state kinetic parameters for the hydrolysis of the colorimetric substrates 2-nitro-5-[(phenylacetyl)amino]-benzoic acid (NIPAB) and D-2-nitro-5-[(phenylglycyl)amino]-benzoic acid (NIPGB) were determined by measuring initial velocities of 5-amino-2-nitro-benzoic acid release ($\Delta\varepsilon_{405nm}$ = 9.09 mM^{-1} cm^{-1}) at 405 nm in a Perkin Elmer Bio40 UV/VIS spectrometer, using substrate concentrations ranging from 5 μM to 10 mM. Data were fitted with the program Origin 6.0 (Microcal Software, Inc.). K_i values for phenylacetic acid and K_m values for non-colorimetric substrates were determined by measuring the inhibition on the hydrolysis of NIPGB as described by Alkema et al. (2002b). The k_{cat} values were determined separately by monitoring the initial velocities of substrate conversion at substrate concentrations of at least $10\times K_m$ by high-performance liquid chromatography (HPLC). Product concentrations were determined at several times in order to obtain at least three data points in the initial phase of conversion. All HPLC analyses were carried out using a 10-cm Chrompack C18 column (5 mm diameter) in connection with Jasco PU-980 pumps and a Jasco MD-910 detector set at 214 nm. Compounds were isocratically

eluted at a flow rate of 1 ml min^{-1} with a solution of 340 mg l^{-1} sodium dodecylsulfate and 680 mg l^{-1} $KH_2PO_4 \cdot 3H_2O$ in a 30:70 (v/v) acetonitrile/water mixture of pH 3.0 (adjusted with diluted phosphoric acid).

Kinetically controlled enzymatic synthesis of β-lactam antibiotics at pH 7.0 was carried out by mixing enzyme with solutions of activated acyl donor [either phenylacetamide (PAA), D-phenylglycine amide (PGA), *p*-hydroxyphenylacetamide (HPAA), or D-*p*-hydroxyphenylglycine amide (HPGA)] and an appropriate β-lactam compound (6-APA or 7-ADCA). The initial concentration of acyl donor was 15 mM in all experiments, whereas the concentration of β-lactam acyl acceptor varied between 1 and 190 mM. All reactants were monitored in time by HPLC analysis and initial rates of formation of the antibiotic (v_{Ps}) and the hydrolyzed acyl donor (v_{Ph}) were determined. Peak areas were related to the concentration of the respective compounds by calibration curves that were established with solutions of the pure compounds. Only for *p*-hydroxypenicillin G, no authentic response factor (mM^{-1}) could be determined due to the lack of a commercially available reference compound. However, as response factors of compounds varying only by the presence of a *p*-hydroxy group (PGA and HPGA, ampicillin and amoxicillin, or cephalexin and cefadroxil, respectively), were found to be very similar (< 10 % difference), we approximated the response given by *p*-hydroxypenicillin G with the one obtained for penicillin G.

2.6. Chemicals

Ampicillin, cefadroxil, and cephalexin were purchased from Sigma; HPAA was from Acros Organics. Penicillin G, amoxicillin, 7-ADCA, 6-APA, PGA, and HPGA were provided by DSM Life Sciences (Delft, The Netherlands). NIPAB and NIPGB were synthesized by reacting phenylacetic acid chloride and D-phenylglycine chloride, respectively, with 5-amino-2-nitro-benzoic acid in a water/acetone mixture. PAA was prepared by adding phenylacetylchloride dropwise to a concentrated ammonia solution, resulting in the formation of a white precipitate, which was filtered off and dried to constant weight. D-Phenylacetyl-L-leucine was obtained through standard organic chemical peptide coupling chemistry. During the synthesis, consecutive washing of the di-protected dipeptide with acid and base ensured that absolutely no free leucine remained in the sample.

3. RESULTS AND DISCUSSION

3.1. Isolation of the new penicillin acylase PAS2

From an enrichment culture for amidase-containing organisms inoculated with sand soil and supplied with PGA as a sole source of nitrogen, genomic DNA was extracted and cloned into a high-copy plasmid vector. Recombinant plasmids were transformed to the leucine-auxotroph *E. coli* host strain TOP10. The clone expressing PAS2 was identified by growth on a selective medium, to which phenylacetyl-L-leucine was added as the only leucine supply.

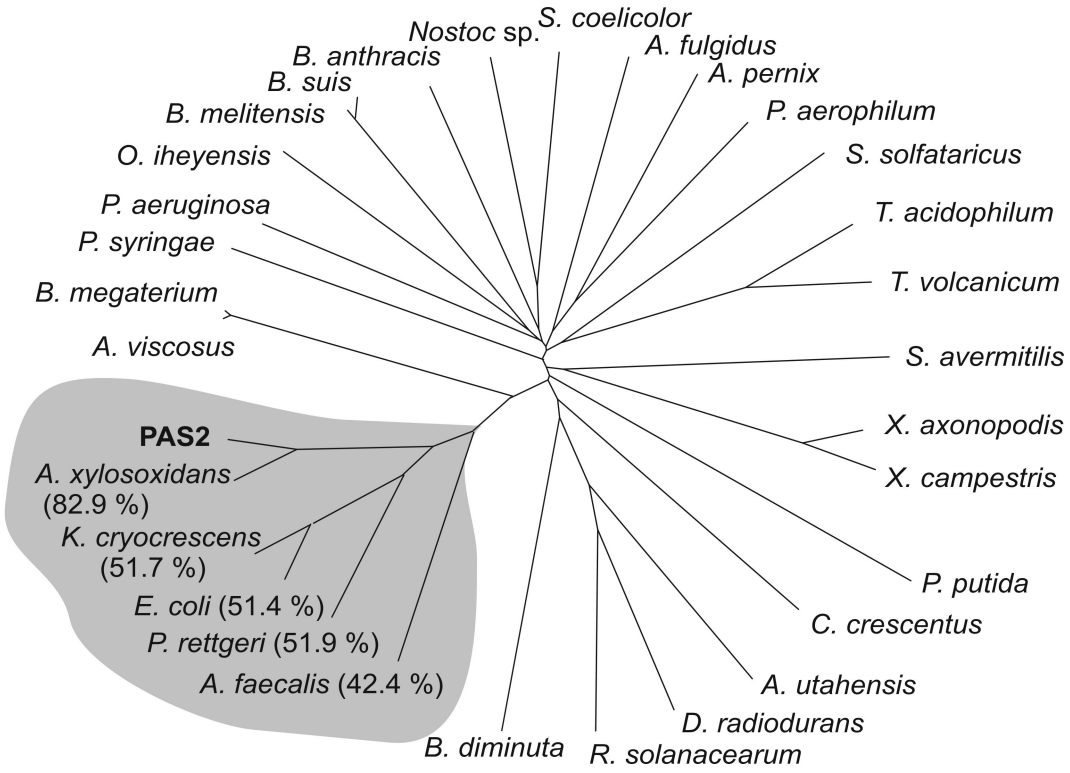


Figure 2. Phylogeny of penicillin acylases. The new enzyme PAS2 falls into a group of closely related enzymes with experimentally verified PA activity (>40 % homology, shaded in gray). Numbers in brackets indicate amino acid sequence identity with PAS2. Sequences identified by BLAST searches (Altschul et al., 1990) with PAS2 as the query were aligned with Megalign software (DNAstar Inc.) and represented in a tree structure with Treeview software (Page, 1996).

The coding sequence of PAS2 was at first identified by sequence analysis of the 7.2 kb insert carried by the original clone and subsequently confirmed by subcloning in the expression vector pBAD/*Myc*-HisA_*Nde*I, which allowed high-level expression of the active protein. A BLAST search (Altschul et al., 1990) with the derived protein sequence yielded a proposed penicillin G acylase of *Achromobacter xylosoxidans* as the strongest hit (82.9 % identity, accession AAP20806) and revealed somewhat lower homology to a group of closely related and well-studied enzymes (Fig. 2), including the PA of *E. coli* ATCC 11105 (51.4 % identity, accession AAA24324). Sequence analysis and comparison with other PAs suggested the expression of PAS2 as a preproprotein, composed of a signal peptide that leads to the translocation of the protein to the periplasm while being cleaved off itself, and two subunits that are separated by a spacer peptide. In *E. coli*, this spacer is removed in the periplasm initiated by an intramolecular autoproteolytic process (Kasche et al., 1999). Mass spectrometric analysis of the purified mature protein confirmed the cleavage of the PAS2 preproprotein between positions 24 and 25, which was also the cleavage site predicted by the PSORT program (Nakai and Horton, 1999) due to the presence of positively charged residues at the N-terminus followed by hydrophobic residues and a consensus pattern for recognition by signal peptidase I (**AXA**) at the C-terminus of the

signal peptide (MKQHLLSAAILAACAGVGAAPA**HA-**QS...). About 30 % of the α-subunit, however, was found to have a 128 Da lower mass. This mass difference corresponds to the molecular weight of a glutamine residue, suggesting that cleavage by the *E. coli* signal peptidase is not very specific and can also occur between positions 25 and 26.

Due to the obtained molecular masses of the small subunit, the exact cleavage point between the α-subunit C-terminus and the spacer could be determined. The removal of the 54 amino acid spacer was found to result in the release of a 229 amino acid α-subunit (25.5 kDa) and a β-subunit of 555 amino acids (61.9 kDa). Hydrolytic activity on the colorimetric penicillin acylase substrate NIPAB was mainly found in periplasmatic extracts and not in the cytosolic or membrane fraction of cells, which supports the idea that PAS2 is processed in a similar way as the *E. coli* PA. The proposed cleavage also yields an N-terminal serine on the β-subunit, which is responsible for the catalytic activity of the enzyme as confirmed by the stoichiometric inactivation of PAS2 by phenylmethylsulfonyl fluoride. Because of the presence of this N-terminal serine that can act as a nucleophile, the observed gene homology and topology, and the activation of the protein by a presumably autocatalytic process, we conclude that PAS2 is a new member of the Ntn-hydrolase superfamily (Brannigan et al., 1995). The characteristic αββα-fold of this class of enzymes was predicted to be present in PAS2 as well when a homology model was made using the structure of *E. coli* PA as a template.

3.2. Biocatalytic performance

The enzymatic hydrolysis of various activated acyl donors, antibiotics, and colorimetric substrates that are typically converted by PAs was studied (Table 1). Due to its primary activity against penicillin G, PAS2 can be classified as a penicillin G acylase (type II PA; Valle et al., 1991) although it can convert a much broader range of β-lactam antibiotics. In general, substrate specificities (k_{cat}/K_m) were found to be similar or higher than for the *E. coli* PA, with comparatively stronger preference of substrates lacking an α-amino substituent.

Compared to the *E. coli* enzyme, PAS2 was found to be much more susceptible towards competitive inhibition by phenylacetic acid (K_i = 14 µM versus 50 µM for *E. coli* PA). Taking into account also its lower specificity for penicillin G, the new enzyme appears to be of limited use for the hydrolytic production of 6-APA from penicillin G where phenylacetic acid is stoichiometrically released. In the hydrolysis of NIPAB, however, PAS2 was found to be about 5-fold more effective, which makes the colorimetric compound PAS2's best substrate tested so far.

3.3. Kinetically controlled antibiotic synthesis

In contrast to hydrolysis, the performance of PAS2 in the synthesis of penicillin G from 15 mM PAA and 25 mM 6-APA at pH 7.0 was found to be significantly better than that of the *E. coli* enzyme, allowing a more than 2-fold higher maximal penicillin G accumulation in the course of reaction (data not shown). Under the same reaction conditions, also 1.2-fold more ampicillin and 1.6-fold more amoxicillin could be

produced with PAS2 (Fig. 3). With PAS2, reactions proceeded at 51 % (ampicillin) and 106 % (amoxicillin) higher pace, measured as the initial rate of antibiotic formation. Because of these promising first results, the new enzyme was studied in more detail, in particular with respect to the production of the semi-synthetic antibiotics ampicillin, amoxicillin, cephalexin, and cefadroxil.

Table 1. Steady-state kinetic parameters of PAS2 and *E. coli* PA for different substrates[a]

	PAS2			*E. coli* PA		
Compound	k_{cat} [s^{-1}]	K_m [mM]	k_{cat}/K_m [mM^{-1} s^{-1}]	k_{cat} [s^{-1}]	K_m [mM]	k_{cat}/K_m [mM^{-1} s^{-1}]
NIPAB	24	0.004	6000	18	0.015	1200
NIPGB	12	0.646	18.6	14	1.3	10.8
PAA	23	0.030	767	46	0.156	295
HPAA	29	0.027	1074	47	0.114	412
PGA	25	12.0	2.1	57	30	1.9
HPGA	16	9.1	1.8	28	12.2	2.3
Penicillin G	25	0.012	2083	39	0.013	3000
Ampicillin	16	0.575	27.8	25	2.5	10.0
Amoxicillin	15	0.399	37.6	17	1.07	15.9
Cephalexin	20	1.3	15.4	29	1.5	19.3
Cefadroxil	13	0.284	45.8	32	0.642	49.8

[a] Coefficients of variation were below 10 % for all data.
NIPAB, 2-nitro-5-[(phenylacetyl)amino]-benzoic acid; NIPGB, D-2-nitro-5-[(phenylglycyl)amino]-benzoic acid; PAA, phenylacetamide; HPAA, *p*-hydroxyphenylacetamide; PGA, D-phenylglycine amide; HPGA, D-*p*-hydroxyphenylglycine amide.

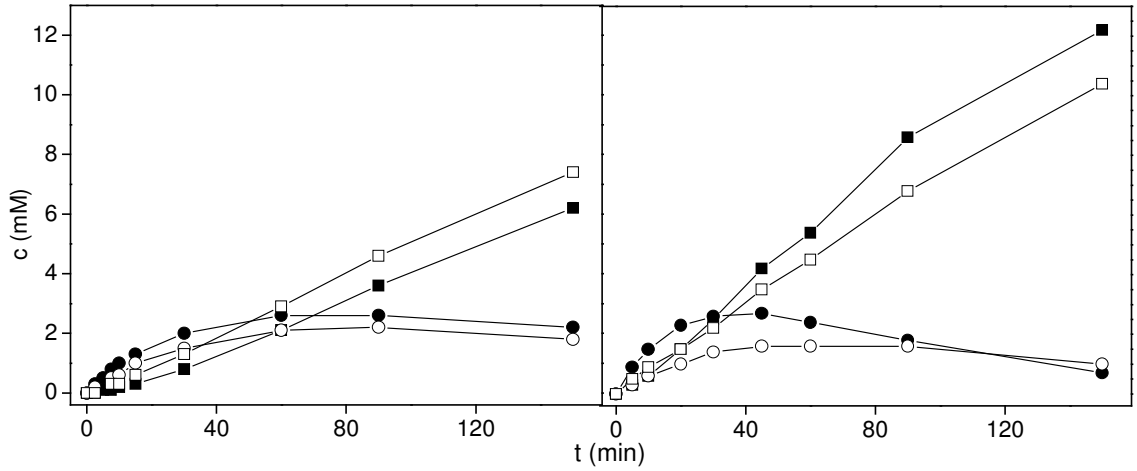


Figure 3. Kinetically controlled synthesis of ampicillin (left) and amoxicillin (right) using 25 mM 6-aminopenicillanic acid (6-APA) and 15 mM D-phenylglycine amide (PGA) or D-*p*-hydroxyphenylglycine amide (HPGA), respectively. PAS2: (●) antibiotic, (■) D-phenylglycine (PG, left) and D-*p*-hydroxyphenylglycine (HPG, right). *E. coli* PA: (○) antibiotic, (□) PG (left) and HPG (right).

Kinetically controlled antibiotic synthesis with PA follows the kinetic scheme represented in Fig. 4. Besides the desired condensation of the activated acyl donor (AD) with the nucleophilic β-lactam compound (N) to the semi-synthetic antibiotic (Ps), PA also catalyzes the nucleophilic attack of water on AD and Ps, leading to the formation of the acyl donor acid as a hydrolytic side-product (Ph). In the beginning of the reaction, however, product hydrolysis can be neglected and the formation of Ph is solely due to hydrolysis of the activated side chain donor. The ratio of the initial rates of antibiotic synthesis (v_{Ps}) and AD hydrolysis (v_{Ph}), the so-called synthesis/hydrolysis ratio, therefore reflects the tendency of the covalent acyl-enzyme intermediate to react with the β-lactam compound instead of with water. The initial synthesis/hydrolysis ratio is given by:

$$\left(\frac{v_{Ps}}{v_{Ph}}\right)_{\text{ini}} = \frac{1}{\gamma} \cdot \frac{[\text{N}]}{\frac{1}{\beta_0 \gamma} + [\text{N}]} \qquad \text{Equation 1}$$

with the complex kinetic parameters β_0 and γ.

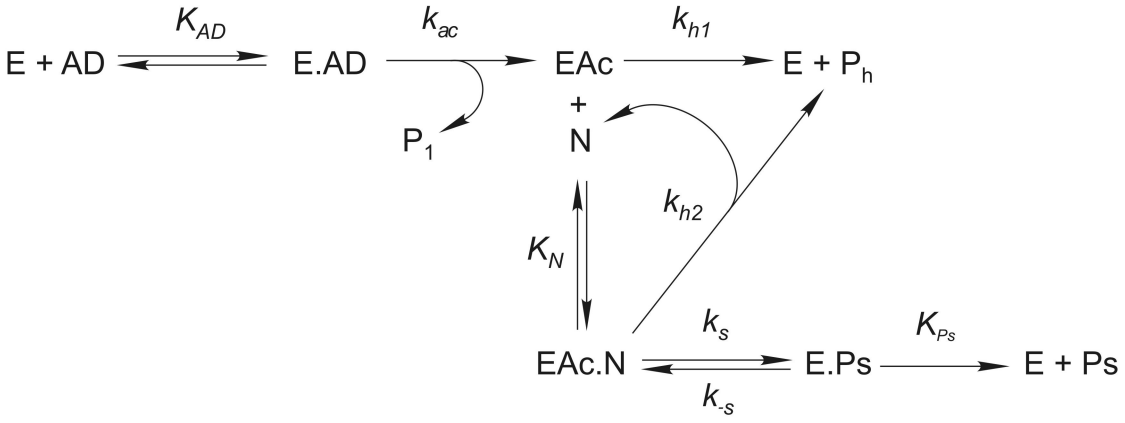


Figure 4. Kinetic scheme of the synthesis and hydrolysis reactions catalyzed by penicillin acylase via the formation of a covalent acyl-enzyme intermediate EAc. Designations: E (free enzyme), AD (activated acyl donor), E.AD (non-covalent enzyme-acyl donor complex), P_1 (first reaction product that is released upon acylation of the enzyme), Ph (product of hydrolysis reactions), N (β-lactam nucleophile), EAc.N (covalent acyl-enzyme intermediate with non-covalently bound nucleophile), Ps (product of the synthesis reaction, i.e. the antibiotic), E.Ps (enzyme-antibiotic complex). Rate constants k_{ac}, k_{h1}, k_{h2}, k_s, and k_{-s} refer to different reaction steps as indicated in the scheme. K_{AD}, K_{Ps} and K_N are the binding constants of AD and Ps to the free enzyme and of N to the acyl-enzyme intermediate, respectively.

From the kinetic scheme, it follows that $\beta_0 = k_s/(k_{h1}K_N)$ and $\gamma = k_{h2}/k_s$ (Youshko et al., 2002a). As can be seen from Eq. 1, the tendency towards the synthesis reaction rises with increasing concentration of the β-lactam compound in a hyperbolic, Michaelis-Menten-type way. Consequently, $1/\gamma$ corresponds to the maximal synthesis/hydrolysis ratio that can be reached under certain reaction conditions and

describes the mode of conversion when the acyl-enzyme intermediate is fully complexed with the nucleophilic β-lactam compound (EAc.N, Fig. 4). At low nucleophile concentrations, when most of the acyl-enzyme intermediate is still free (EAc), the synthesis/hydrolysis ratio linearly increases with the nucleophile concentration according to:

$$\left(\frac{v_{Ps}}{v_{Ph}}\right)_{\text{ini}} = \beta_0 \cdot [\text{N}] \qquad \text{Equation 2}$$

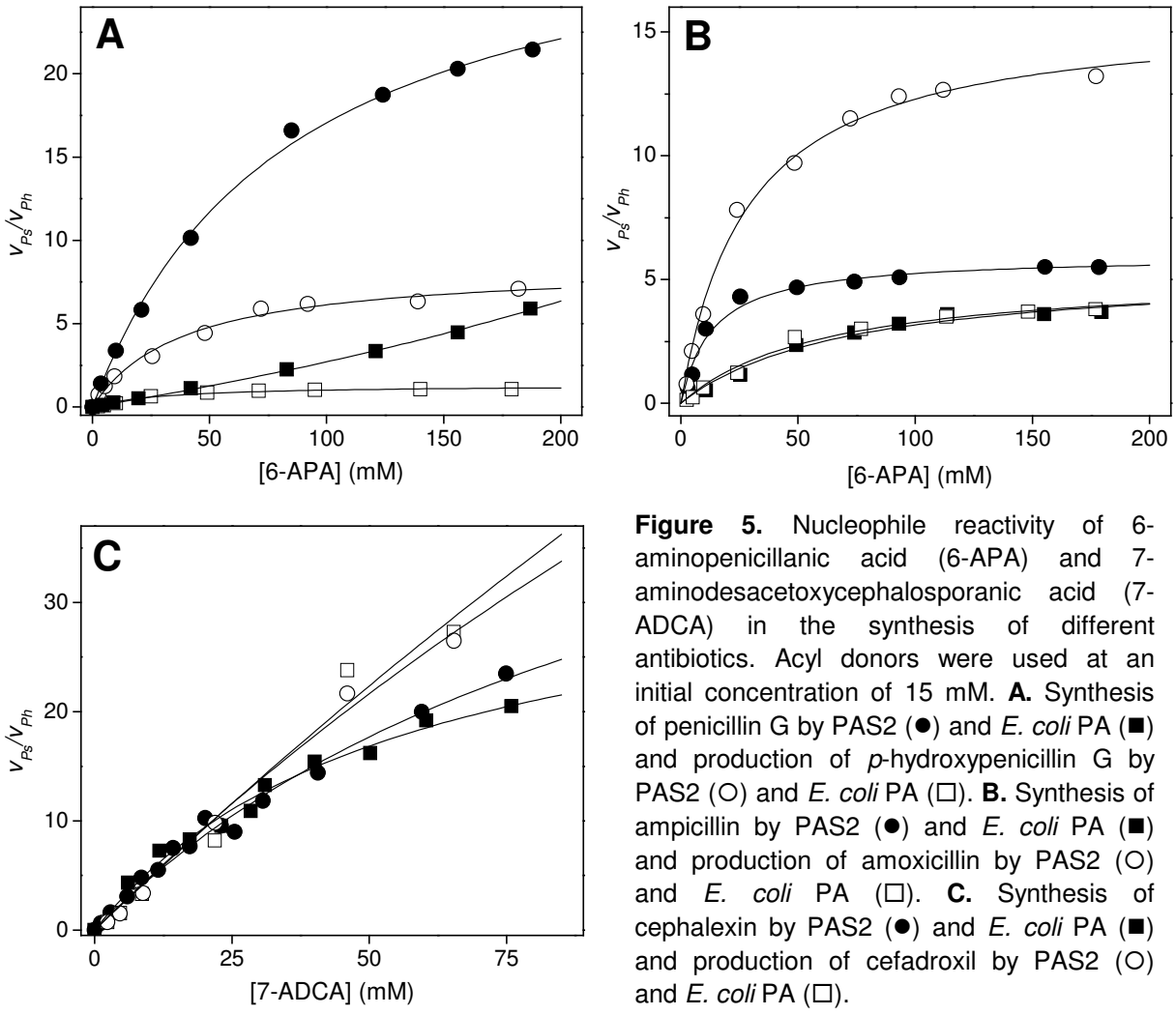


Figure 5. Nucleophile reactivity of 6-aminopenicillanic acid (6-APA) and 7-aminodesacetoxycephalosporanic acid (7-ADCA) in the synthesis of different antibiotics. Acyl donors were used at an initial concentration of 15 mM. **A.** Synthesis of penicillin G by PAS2 (●) and *E. coli* PA (■) and production of *p*-hydroxypenicillin G by PAS2 (○) and *E. coli* PA (□). **B.** Synthesis of ampicillin by PAS2 (●) and *E. coli* PA (■) and production of amoxicillin by PAS2 (○) and *E. coli* PA (□). **C.** Synthesis of cephalexin by PAS2 (●) and *E. coli* PA (■) and production of cefadroxil by PAS2 (○) and *E. coli* PA (□).

We determined initial synthesis/hydrolysis rate ratios in the synthesis of a number of antibiotics, using a wide range of different β-lactam nucleus concentrations (Fig. 5). In the resulting $(v_{Ps}/v_{Ph})_{\text{ini}}$ versus [N] plots, β_0 constitutes the initial slope, whereas $1/\gamma$ is the asymptotic value to which the curve saturates at high β-lactam nucleophile concentrations. PAS2 reached clearly higher synthesis/hydrolysis ratios over the whole concentration range for all 6-APA derived antibiotics tested. In the synthesis of penicillin G by *E. coli* PA and cefadroxil by either of the two studied enzymes, no

saturation was observed even at the highest feasible β-lactam concentrations, compromising the determination of γ. In contrast, β_0 could be readily determined for all antibiotics and turned out to be significantly higher for 6-APA in PAS2- than in *E. coli* PA-catalyzed reactions (Fig. 5, A and B). Remarkably, 7-ADCA reactivity was almost identical for both studied enzymes irrespective of the kind of acyl donor used (Fig. 5.C).

To fully describe the catalytic behavior of penicillin acylase, a third parameter, α, is required. This parameter delimits the maximal amount of antibiotic that can be accumulated in the course of reaction and describes the susceptibility of the enzyme to hydrolyze the initially formed product. Basically, α quantifies the enzyme's preference to hydrolysis of the antibiotic over hydrolysis of the activated acyl donor:

$$\alpha = \frac{\left(\frac{k_{cat}}{K_m}\right)_{\mathrm{Ps}}}{\left(\frac{k_{cat}}{K_m}\right)_{\mathrm{AD}}} \qquad \text{Equation 3}$$

To allow high product accumulation, the specificity for the acyl donor should be as high as possible with a low reactivity towards the desired synthesis product, i.e. α should be small. Table 2 summarizes the complex kinetic parameters determined for PAS2 and the *E. coli* PA.

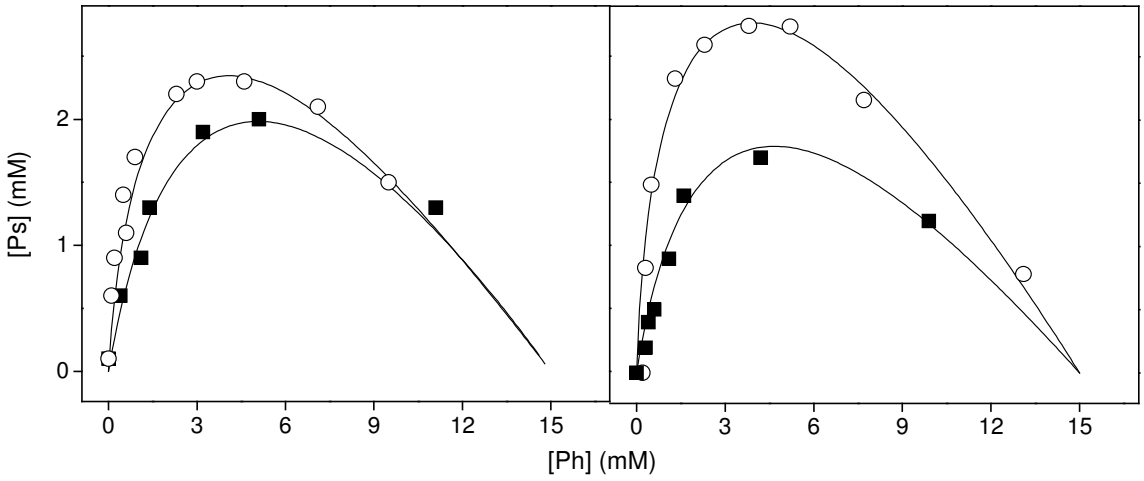


Figure 6. Product formation in the course of ampicillin (left) and amoxicillin synthesis (right), using PAS2 (○) and *E. coli* PA (■). Ps is the synthetic product, i.e. the antibiotic, and Ph is the hydrolyzed acyl donor, i.e. D-phenylglycine (PG, left) or D-*p*-hydroxyphenylglycine (HPG, right). All reactions were carried out with 15 mM acyl donor and 25 mM β-lactam nucleophile. Solid lines were calculated by numerical integration of Eq. 4, using the experimentally determined parameters given in Table 2 and substrate concentrations indicated above.

Table 2. Complex kinetic parameters characterizing the efficiencies of PAS2 and *E. coli* PA in antibiotic synthesis

Acyl donor	β-Lactam nucleophile	Antibiotic (P_s)	β_0 (M^{-1})		$1/\gamma$ [a]	
			PAS2	*E. coli* PA	PAS2	*E. coli* PA
PAA	6-APA	Penicillin G	373	30	31	n.d.
HPAA	6-APA	*p*-Hydroxypenicillin G	204	35	9	1.5
PGA	6-APA	Ampicillin	448	76	6	5
HPGA	6-APA	Amoxicillin	560	86	16	5
PGA	7-ADCA	Cephalexin	509	638	59	36
HPGA	7-ADCA	Cefadroxil	444	427	n.d.	n.d.

Acyl donor	β-Lactam nucleophile	Antibiotic (P_s)	α		Ps_{max} (mM) [b]	
			PAS2	*E. coli* PA	PAS2	*E. coli* PA
PAA	6-APA	Penicillin G	2.7	10.2	1.8	0.8
HPAA	6-APA	*p*-Hydroxypenicillin G	n.d.	n.d.	n.d.	n.d.
PGA	6-APA	Ampicillin	13.2	5.3	2.4	2.0
HPGA	6-APA	Amoxicillin	20.9	6.9	2.7	1.7
PGA	7-ADCA	Cephalexin	7.3	10.2	4.1	3.7
HPGA	7-ADCA	Cefadroxil	25.4	21.7	2.4	2.2

[a] $1/\gamma$ corresponds to the maximal initial synthesis/hydrolysis ratio that can be reached in the synthesis of a given antibiotic.

[b] Maximal level of accumulation of the synthesis product obtained with 25 mM of β-lactam nucleophile and 15 mM of acyl donor.

n.d., not determined; PAA, phenylacetamide; HPAA, *p*-hydroxyphenylacetamide; PGA, D-phenylglycine amide; HPGA, D-*p*-hydroxyphenylglycine amide; 6-APA, 6-aminopenicillanic acid; 7-ADCA, 7-aminodesacetoxycephalosporanic acid

As observed for the nucleophile reactivity, differences in α between the two studied enzymes were more clearly expressed for 6-APA than for 7-ADCA-derived antibiotics, which suggests that the binding mode of 6-APA in the active site is altered in PAS2 while being similar to that in *E. coli* PA for 7-ADCA (Table 2). However, amino acid residues that have previously been identified to be important in β-lactam binding in the *E. coli* enzyme (Alkema et al., 2002b and 2000) are conserved in PAS2. Consequently, only subtle differences in the orientation of these residues or additional and not yet identified residues are expected to be responsible for the observed deviations in binding of the β-lactam nucleophile. Besides altered binding of the nucleophilic β-lactam group, binding of the acyl donor is also different in PAS2, which is most obvious for phenylacetamide. For this compound, a 5-fold higher apparent affinity was observed in PAS2, leading to a strongly improved α parameter for the new enzyme. Together with the high reactivity of 6-APA, this fact allowed the accumulation of 2.3 times more penicillin G than with *E. coli* PA.

Experimental data of product concentrations ([Ps] and [Ph]) in the course of the reaction were in good agreement with the theoretically expected ones calculated on the basis of the kinetic model (Eq. 4) as is exemplified for ampicillin and amoxicillin in Fig. 6. These results show that the model, which was originally developed for *E. coli* PA, also applies to PAS2.

$$\frac{d[Ps]}{d[Ph]} = \frac{\beta_0[N][AD] - \alpha[Ps](1+\beta_0\gamma[N])}{(1+\beta_0\gamma[N])([AD]+\alpha[Ps])}, \qquad \text{Equation 4}$$

with $[AD]_0 = [AD] + [Ps] + [Ph]$ and $[N]_0 = [N] + [Ps]$

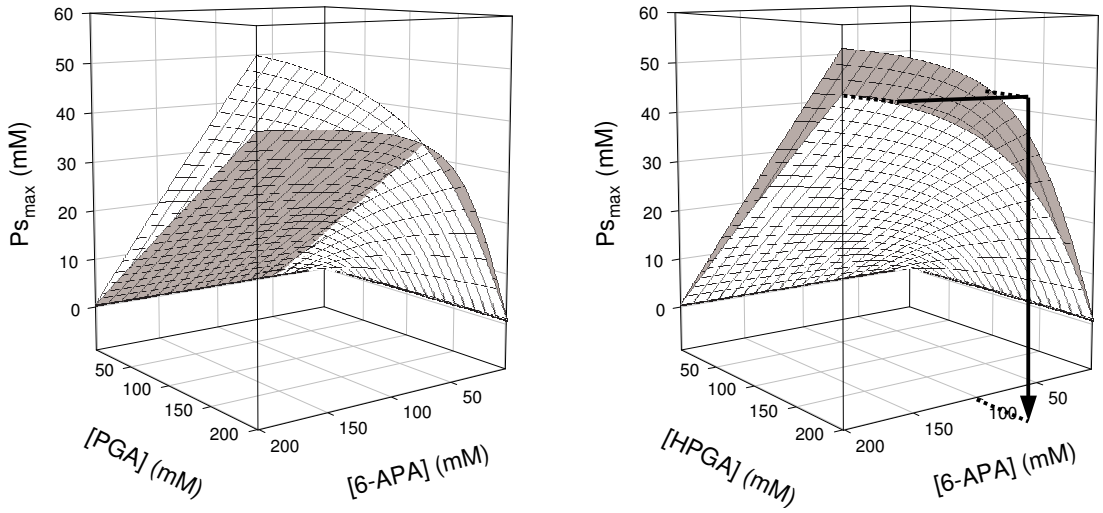


Figure 7. Maximal level of accumulation (Ps_{max}) of ampicillin (left) and amoxicillin (right) predicted for different substrate concentrations with PAS2 (gray) and *E. coli* PA (white). The arrow indicates the concentration of 6-aminopenicillanic acid (6-APA) that allows the accumulation of the same amount of amoxicillin as maximally obtained with *E. coli* PA from 200 mM 6-APA. PGA, D-phenylglycine amide; HPGA, D-*p*-hydroxyphenylglycine amide.

By using this model, product yields can also be calculated for higher substrate concentrations, which may be more relevant for industrial processes. As can be seen from Fig. 7, the relative performance of PAS2 in ampicillin as well as amoxicillin synthesis is most clearly improved compared to the *E. coli* PA in the low 6-APA region (<100 mM) due to the significantly higher β_0 parameters of PAS2. When higher concentrations of both reactants are used, however, relative differences in maximal antibiotic yield decrease and for ampicillin, the accumulation level is even somewhat lower than with the *E. coli* enzyme. For the production of amoxicillin, in contrast, PAS2 is more effective over the whole range of substrate concentrations. Relative improvements range from 240 % in the low PGA/low 6-APA region to about 14 % at high substrate concentrations. This increase in synthetic performance would allow the use of lower concentrations of substrates, particularly 6-APA, in the production

process. To reach a 48.5 mM amoxicillin concentration, for instance, which is the maximum achievable with *E. coli* PA under the modeled conditions, only 100 mM instead of 200 mM 6-APA would be required with PAS2 at a PGA concentration of 200 mM.

In conclusion, high specific activity combined with improved kinetic properties that allow increased levels of antibiotic accumulation constitute the main advantages of the newly isolated enzyme in the kinetically controlled production of semi-synthetic penicillins, particularly of amoxicillin. The observed high turnover rates are especially remarkable since improved synthetic capacity has been found to be coupled to a loss of enzyme activity for site-directed mutants of *E. coli* PA (Alkema et al., 2000; Alkema et al., 2002b). The further improvement of PAS2 by mutagenesis techniques is described in *Chapter 6*.

The discovery of PAS2 demonstrates once more that the steadily rising request for improved biocatalysts may in part be satisfied by screening the almost untapped environmental gene pool.

ACKNOWLEDGEMENTS

We thank Dr. H.P. Permentier for mass spectrometric analysis of PAS2.

6

Increasing the synthetic performance of penicillin acylase PAS2 by structure-inspired semi-random mutagenesis

Esther M. Gabor and Dick B. Janssen

Submitted

A semi-random mutagenesis approach was followed to increase the performance of penicillin acylase PAS2 in the kinetically controlled synthesis of ampicillin from 6-aminopenicillanic acid (6-APA) and activated D-phenylglycine derivatives. We directed changes in amino acid residues to positions close to the active site that are expected to affect the catalytic performance of penicillin acylase: αR160, αF161, and βF24. From the resulting triple mutant gene bank, 6 improved PAS2 mutants were recovered by only screening 700 active mutants with a HPLC-based screening method. A detailed kinetic analysis of the three most promising mutants, T23, TM33, and TM38, is presented. These mutants allowed the accumulation of ampicillin at 4 to 5 times higher concentrations than the wild-type enzyme, using D-phenylglycine methyl ester as the acyl donor. At the same time, the loss of activated acyl donor due to the competitive hydrolytic side reactions could be reduced to less than 20 % with the mutant enzymes compared to > 80 % with wild-type PAS2. Although catalytic activity dropped by a factor 5 to 10, the enhanced synthetic performance of the recovered penicillin acylase variants makes them interesting biocatalysts for the production of β-lactam antibiotics.

1. INTRODUCTION

Already since several decades, penicillin acylase (penicillin amidase, PA, EC 3.5.1.11) is used for the hydrolytic release of 6-aminopenicillanic acid (6-APA) from fermentatively obtained natural penicillins, which is a central step in the production of semi-synthetic β-lactam antibiotics. Recently, also biocatalytic processes using this enzyme for the kinetically controlled condensation of activated acyl moieties with 6-APA or other β-lactam nuclei have been developed (Bruggink et al., 1998). Although being competitive with traditional chemical synthesis routes, these enzymatic procedures suffer from the fact that the desired synthesis of β-lactam antibiotics is accompanied by two side reactions, leading to the loss of activated acyl donor and limiting the product yield: (1) hydrolysis of the activated acyl donor, and (2) hydrolysis of the formed antibiotic. Efforts to overcome this problem have mostly focused on a change in reaction conditions, such as pH optimization (Youshko et al., 2002b), the use of supersaturated substrate solutions (Youshko et al., 2004) or immobilized enzyme preparations (Alvaro et al., 1990) and medium engineering (Fernández-Lafuente et al., 1996). However, it was shown that the intrinsic kinetic properties of the employed biocatalyst also have a major impact on the efficiency of the synthetic reaction (Hernández-Justiz et al., 1999; Youshko et al., 2002a; Alkema et al., 2003).

During PA-catalyzed antibiotic synthesis, the formation of the synthetic product Ps, i.e. the antibiotic, as well as the formation of the hydrolytic byproduct Ph proceed via the formation of a covalent acyl-enzyme intermediate EAc, which is the rate-limiting step in the catalytic process (kinetic scheme: see *Chapter 5*, Fig. 4). EAc can then be attacked by one of two possible nucleophiles: either the β-lactam nucleus N, leading to the desired product Ps, or by a water molecule, which results in the

formation of Ph. The ratio between the initial production rates of synthetic product (v_{Ps}) and hydrolytic product (v_{Ph}), and the maximal level of product accumulation that is obtained (Ps_{max}) depend on the initial substrate concentrations $[N]_0$ and $[AD]_0$, and the kinetic parameters of the enzyme. In fact, antibiotic formation in PA-catalyzed reactions is sufficiently described by only three complex kinetic parameters, α, β_0, and γ, according to Eq. 1 (Youshko et al., 2002a):

$$\frac{d[Ps]}{d[Ph]} = \frac{\beta_0[N][AD] - \alpha[Ps](1+\beta_0\gamma[N])}{(1+\beta_0\gamma[N])([AD]+\alpha[Ps])}, \qquad \text{Equation 1}$$

with $[AD]_0 = [AD] + [Ps] + [Ph]$ and $[N]_0 = [N] + [Ps]$

Experimentally, parameters β_0 and γ can be directly determined, since they relate the initial rates of product and byproduct formation to the concentration of nucleophilic β-lactam compound in a Michaelis-Menten-type equation:

$$\left(\frac{v_{Ps}}{v_{Ph}}\right)_{ini} = \frac{1}{\gamma} \cdot \frac{[N]}{\frac{1}{\beta_0\gamma} + [N]} \qquad \text{Equation 2}$$

Here, $1/\gamma$ is the maximum that is reached for $[N] \rightarrow \infty$, while β_0 constitutes the initial slope of a $(v_{Ps}/v_{Ph})_{ini}$ vs. [N] plot. The third parameter, α, introduces the fact that the synthesized antibiotic is also a substrate for the enzyme, leading to its hydrolysis.

$$\alpha = \frac{\left(\frac{k_{cat}}{K_m}\right)_{Ps}}{\left(\frac{k_{cat}}{K_m}\right)_{AD}} \qquad \text{Equation 3}$$

Parameter α actually reflects the relative specificity of PA for the antibiotic with respect to the activated acyl donor and should be as small as possible to allow high product accumulation in a synthesis reaction. Parameters β_0 and $1/\gamma$, in contrast, preferably should be high to prevent the formation of large amounts of hydrolytic byproduct. The use of a PA with favorable kinetic parameters for the synthesis of a specific antibiotic thus takes a key position in the design of an efficient production process.

Biocatalysts with desired properties can be either obtained from nature, applying search strategies that range from traditional screening of culture collections to screening of metagenomic gene banks (Handelsman et al., 1998; Lorenz et al., 2002), or by improving already available enzymes (Kazlauskas, 2000; Zhao et al., 2002). Examples for the first approach are the PAs of *Kluyvera citrophila*, *Bacillus megaterium*, *Proteus rettgeri*, or *Alcaligenes faecalis* that were all detected for the first time by assaying pure bacterial cultures (Huber et al., 1972; Barbero et al., 1986). Unfortunately, none of these enzymes seems to be better suited for synthetic purposes than the already known PA of *E. coli* (Hernández-Justiz et al., 1999). In contrast, an

enzyme with better kinetic properties for 6-APA-derived antibiotic production, PAS2, was isolated from a soil metagenomic gene library (*Chapter 4*). PAS2 allows significantly improved antibiotic yields and, concomitantly, lower amounts of hydrolytic byproduct than the industrially employed *E. coli* PA (*Chapter 5*), with which it has 51.4 % of its amino acid residues in common.

In this study, we explored the possibility to further increase the synthetic performance of PAS2 for ampicillin production, using a semi-random mutagenesis procedure. By simultaneous randomization of three amino acid residues close to the active site, a gene library comprising 6,000 original mutants was created. Since the three selected residues are conserved in the *E. coli* PA and are known to greatly affect the kinetic properties of this enzyme, we anticipated that the mutants would show significant variations in catalytic behavior. Indeed, 6 clearly improved enzymes were found when screening only 700 active mutants. A detailed kinetic analysis of the three most promising PAS2-derivatives is presented.

2. MATERIALS AND METHODS

2.1. Chemicals

Ampicillin and cephalexin were purchased from Sigma. D-Phenylglycine (PG) was from Acros. 6-Aminopenicillanic acid (6-APA), 7-aminodesacetoxycephalosporanic acid (7-ADCA), D-phenylglycine methyl ester (PGM), and D-phenylglycine amide (PGA) were provided by DSM Life Sciences (Delft, The Netherlands). The colorimetric substrate 2-nitro-5-[(phenylacetyl)amino]-benzoic acid (NIPAB) was synthesized by reacting phenylacetic acid chloride with 5-amino-2-nitro-benzoic acid in a water/acetone mixture. D-phenylacetyl-L-leucine (PAL) was obtained through standard organic chemical peptide coupling chemistry. Care was taken that no free leucine was present in the prepared sample.

2.2. Plasmids and strains

For all plasmid constructs described in this study, *E. coli* TOP 10 [Δ(ara-leu)7697, Invitrogen] was used as the host organism. pBADPAS2 is a 6.7 kb plasmid carrying the wild-type *pas2* gene under the control of the pBAD promoter (*Chapter 5*). pPAS2CAT is a similar construct, just conferring chloramphenicol instead of ampicillin resistance to prevent interference of β-lactamase activity with β-lactam antibiotic synthesis experiments in whole cells. To obtain this construct, pBADPAS2 was cut with *Bsp*HI (New England Biolabs) to remove the β-lactamase gene. The remaining 5.7 kb fragment was isolated from gel with the QIAEX II gel extraction kit (Qiagen), blunted with Klenow polymerase (Invitrogen) and ligated to an also gel-purified blunt 1.2 kb fragment with T4 ligase (Invitrogen). The 1.2 kb fragment carried a chloramphenicol acetyltransferase gene (*cat*) and originated from the plasmid pEC (provided by DSM, Delft, The Netherlands), from which it was removed by *Bsp*HI/*Eco*RI restriction (New England Biolabs). All DNA manipulations were carried out according to the instructions of the manufacturers of the used materials. The

construct was transformed to *E. coli* by electroporation and confirmed by restriction analysis. All media used to grow clones carrying the wild-type pPAS2CAT plasmid or mutant derivatives were supplied with 68 mg l^{-1} chloramphenicol.

2.3. Triple site-saturating mutagenesis

Two degenerate primers were designed, one for the random mutagenesis of amino acid residues αR160 and αF161 (5'gtg ggc ac**t** atg gcc aac nng/c nng/c tcg gac gcc aat agc gaa atc g 3'), and one for position βF24 (5'ggc ggg gtt cca cca gcc c/gnn ctg cgg gcc gtt **ta**a cag 3'). For the easy identification of real mutants, primer αR160X/αF161X was set up to remove a unique *Nco*I restriction site and primer βF24X to introduce an additional *Mse*I site, both leading to characteristic differences in the restriction patterns of wild-type and mutant plasmids (restriction sites underlined, mutations in bold). To reduce codon bias, the third position of each random codon was fixed to be a guanine or a cytosine. An outline of the used mutagenesis procedure is shown in Fig. 1. In the first step, a 500-bp fragment containing all three randomized codons was PCR-amplified with the two degenerate primers described above. The reaction was carried out with *Pfu Turbo* polymerase (Stratagene) under standard conditions, using 50 ng pPAS2CAT as the template and 125 ng of each of the primers. The amplified fragment (megaprimer) was purified on gel with the QIAEX II gel extraction kit (Qiagen) and used as a forward and reverse primer set in a normal QuickChange site-directed mutagenesis procedure (Stratagene) according to the instructions of the manufacturer. After ethanol precipitation, the reaction mixture was resuspended in water and mutant plasmids were transformed to *E. coli* TOP10 by electroporation. Transformants were plated on LB agar (Sambrook et al., 1989). Mutant strains obtained by this procedure usually carry two plasmid populations with different mutations, one originating from each of the different strands of the original (chimeric) mutant plasmid. To solve this problem, all colonies were washed from the agar plates and used for plasmid

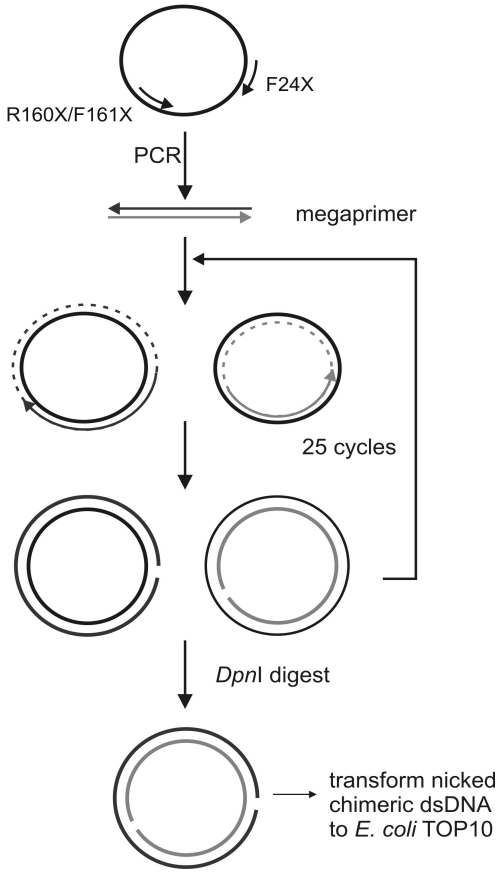


Figure 1. Mutagenesis procedure used in this study. In the first step, a PCR product is prepared, using two degenerate primers that carry the desired randomized codons and the wild-type plasmid as the template. This product is then used as a megaprimer (3 randomized positions) in a standard QuickChange site-directed mutagenesis procedure (Stratagene) to yield the desired triple mutants.

preparation with the High Pure Plasmid Isolation Kit (Roche). The plasmid mix was then retransformed to *E. coli* to produce transformants, each carrying only one type of plasmid. The resulting amplified gene bank was then subjected to growth selection and screening. Interesting mutants were sequenced with at least 2-times coverage of each base by BaseClear Holding BV (Leiden, The Netherlands).

2.4. Selection of active mutants and screening for improved synthesis

Active mutants were selected by their ability to grow on phenylactyl-L-leucine as the sole source of leucine (Forney and Wong, 1989) in a minimal medium supplied with 0.2 % (w/v) glucose and 15 g l^{-1} agarose as described before (*Chapter 4*). PAL was added at a concentration of 10 mg l^{-1} and 0.2 % (w/v) arabinose were included in the medium to induce protein expression from P_{BAD}. After 5 days of growth at 30°C, single colonies were transferred from the agar plates to 96-well microtiter plates (MTPs) filled with 200 µl liquid LB medium per well. After overnight incubation at 30°C and 200 rpm orbital shaking, 50 µl of 50 % glycerol were added per well and plates were stored as frozen stocks at –80°C.

For the first round of screening by high-performance liquid chromatography (HPLC), plates were defrosted and 20 µl aliquots of the stock cell suspensions were transferred to fresh MTPs and used to inoculate 180 µl of LB medium. MTPs were incubated at 17°C with orbital shaking at 200 rpm for 48 h before 50 µl of 10 % arabinose solution were added. After incubation for another 24 h at 17°C, cells were collected by centrifugation for 5 min at 2,500 rpm in a MSE Mistral 2000 MTP centrifuge. They were resuspended in 200 µl of substrate solution per well (15 mM PGA and 10 mM 6-APA in 50 mM potassium phosphate buffer, pH 7.0) and incubated at 30°C. After 24 h, 10 µl of the reaction mixtures were transferred to fresh MTPs and quenched by the addition of 90 µl HPLC eluens [340 mg l^{-1} sodium dodecylsulfate and 680 mg l^{-1} $KH_2PO_4 \cdot 3H_2O$ in a 30:70 (v/v) acetonitrile/water mixture of pH 3.0 (adjusted with diluted phosphoric acid)]. Liquid handling was done with a Plato 3001 automated pipetting station (Rosys AG, Switzerland) that also served as a 96-well MTP format autosampler for HPLC injection. HPLC analyses were carried out using a 3-cm Alltech Alltime C18 3u column in connection with Jasco PU-980 pumps and a Jasco MD-910 detector set at 214 nm. Compounds were isocratically eluted at a flow rate of 5 ml min^{-1} with HPLC eluens, which allowed the analysis of one sample in 3 min. The ratio of the areas of the product peaks, i.e. ampicillin and PG, in each chromatogram was determined and compared to the one of the *E. coli* (pPAS2CAT) control. Mutant strains exhibiting an at least 2-fold increased ampicillin/PG peak ratios or showing exceptionally high conversion rates were analyzed in more detail for their performance in the synthesis of ampicillin.

In this second round of HPLC screening, complete progress curves for the synthesis of ampicillin from 15 mM PGA and 25 mM 6-APA in 50 mM potassium phosphate buffer (pH 7.0) were recorded at 30°C, using periplasmatic extracts and enzyme concentrations of 5-10 µM. Extracts were prepared and analyzed for their content of active penicillin acylase as described in paragraph 2.5. A similar HPLC setup as for the robotic screening served for this analysis, however, based on a 10-cm Chrompack C18 column. Here, compounds were isocratically eluted at a flow rate of 1

ml min^{-1}. Peak areas were related to the concentration of the respective compounds by calibration curves that were established with solutions of the pure compounds.

2.5. Preparation of periplasmatic extracts and enzyme purification

Periplasmatic extracts were prepared at different scales. For the verification of whole cell screening results in the second round of HPLC screening, 5-ml cultures were used as a starting material, while 1-l cultures were used for protein purification. In both cases, clones containing the wild-type or mutant PAS2 gene were grown in LB at 17°C with rotary shaking at 200 rpm. After 2 days of growth, the medium was supplied with 0.8 % arabinose for the induction of protein expression. After another 24 h of incubation, cells were harvested by centrifugation at 5,000 g for 10 min (4°C). To prepare a periplasmatic extract, cells were resuspended in 1/10 of the original culture volume of ice-cold osmotic shock buffer (20 % sucrose, 100 mM Tris-HCl, 10 mM EDTA; pH 8.0) and centrifuged as described above. Cell walls were disrupted by resuspending the cell pellet in 1/50 of the original culture volume of ice-cold 1 mM EDTA and cell debris was removed by another centrifugation step.

Wild-type PAS2 and mutant enzymes were purified as described before (*Chapter 5*). The amount of active enzyme in penicillin acylase preparations was determined by titration with the irreversible inhibitor phenylmethylsulfonyl fluoride (PMSF). Enzyme preparations were incubated with different concentrations of PMSF in 50 mM potassium phosphate buffer (pH 7.0) at room temperature for 15 min. Residual activity was measured with NIPAB or ampicillin as a substrate. The release of 5-amino-2-nitro-benzoic acid from NIPAB was followed spectrophotometrically at 405 nm ($\Delta\varepsilon_{405nm}$ = 9.09 mM^{-1} cm^{-1}) with a Perkin Elmer Bio40 UV/VIS spectrometer, while hydrolysis of ampicillin was monitored by withdrawing samples from the reaction mixtures at different time points in the initial phase of conversion and analyzing them by HPLC (10-cm Chrompack C18 column).

2.6. Determination of kinetic parameters

All enzymatic conversions were carried out in 50 mM potassium phosphate buffer (pH 7.0) at 30°C. Steady-state kinetic parameters of the mutant enzymes for the hydrolysis of PGA, PGM, ampicillin, and cephalexin were determined by monitoring the initial velocities of substrate conversion at various substrate concentrations by HPLC (10-cm Chrompack C18 column). Product concentrations were determined at several times in order to obtain at least three data points in the initial phase of conversion. Steady-state parameters of PAS2 were determined as described before (*Chapter 5*).

Kinetically controlled enzymatic synthesis of ampicillin and cephalexin was carried out by mixing enzyme with solutions of activated acyl donor (PGA or PGM) and 6-APA or 7-ADCA, respectively. The initial concentration of acyl donor was 15 mM in all experiments, whereas the concentration of β-lactam acyl acceptor varied between 0.5 and 200 mM. All reactants were monitored in time by HPLC analysis (10-cm Chrompack C18 column) and initial rates of formation of the antibiotic (v_{Ps}) and the hydrolyzed acyl donor (v_{Ph}) were determined.

3. RESULTS AND DISCUSSION

3.1. Structure-inspired semi-random mutagenesis

Penicillin acylase PAS2 (accession AY573298) and *E. coli* PA (accession AAA24324) share 51.4 % of their amino acid residues. This degree of sequence identity was sufficiently high to obtain a homology model of PAS2 based on the known 3-D structure of the *E. coli* enzyme (Duggleby et al., 1995). The model provided by the Swiss-Model service (http://www.expasy.org/swissmod/SWISS-MODEL.html) did not reveal significant changes of the geometry of the active site, and the spatial orientation of several amino acid residues that were shown to significantly affect the catalytic behavior of *E. coli* PA (Alkema, 2002a) perfectly matched between the two proteins. This agreement as well as the conservation in a number of different other PAs, such as of *Proteus rettgeri* (accession A56681), *Alcaligenes faecalis* (accession AAD11517) or *Kluyvera citrophila* (A26528), prompted us to chose the three residues αR160, αF161, and βF24 of PAS2 (corresponding to positions α145, α146, and β24 in *E. coli* PA) as targets for mutagenesis. Alkema et al. (2000) showed that residues αR145 and αF146 have an important impact on the affinity of the *E. coli* enzyme for phenylacetylated substrates and the positioning of substrates towards the catalytic serine in the active site. Via an induced fit mechanism, they participate in the formation of the β-lactam binding site. When 6-APA or derived antibiotics are bound, both residues interact with the β-lactam moiety: αR145 with the negatively charged carboxylate group via two bridging water molecules, and αF146 by van der Waals interactions with the 2β-methyl group of the thiazolidine ring (Brannigan et al., 2000; Done et al., 1998). The phenyl ring of the side-chain of αF146 also participates in the formation of the AD binding pocket by interacting with the phenyl ring of the acyl compound. On the other side of the binding pocket, another phenylalanine is located, βF24, which also shows hydrophobic interaction with the acyl donor (Alkema et al., 2002b).

While the importance of the three residues in *E. coli* PA – and most likely also in PAS2 - is obvious, it is though difficult to predict what effect a specific mutation will have on the catalytic performance of the enzyme in a kinetically controlled synthesis reaction. This is due to the fact that both the activated acyl donor as well as the antibiotic can serve for the formation of the covalent acyl-enzyme intermediate (EAc). Therefore, optimization of e.g. the enzyme acylation reaction in order to increase the reactivity with the acyl donor usually also results in improved hydrolysis of the formed antibiotic. Also engineering the enzyme with respect to improved deacylation by the β-lactam nucleophile is not straightforward, since increased reactivity of N is often accompanied by an increased specificity also for the β-lactam antibiotic. In the end, only subtle changes in the geometry of the acyl-enzyme or the orientation and activation of the deacylating nucleophiles cause altered rates of product and byproduct formation and influence Ps_{max}. These changes are mostly difficult to rationalize, which is particularly true for PAS2, of which an experimental 3-D structure is not yet available. Therefore, we chose to randomize the respective amino acid residues rather than to follow a completely rational approach. By simultaneously changing residues at

three different sites, we were able to create a gene bank of high sequence diversity, while focusing on mutants that are likely to show significantly altered catalytic performance.

The procedure used to create the triple mutant bank of PAS2 is illustrated in Fig. 1. Since two of the target positions for randomization concern adjacent amino acid residues, we only needed to design two degenerate primers: one for αR160 and αF161, and another for residue βF24. With these two primers, a (double-strand) megaprimer was prepared by simple PCR amplification that could subsequently be used in a standard QuickChange mutagenesis procedure. In this way, 6,000 original clones were created that were stored as an amplified gene library and screened for improved ampicillin synthesis.

3.2. Screening for higher ampicillin yields

The development of a direct screening method for increased antibiotic yields, e.g. by HPLC analysis, is difficult due to the temporary nature of product accumulation. After reaching a maximum (Ps_{max}), the product concentration decreases due to the enzymatically catalyzed hydrolysis of the antibiotic. Earlier experiments showed that an increase in $(v_{Ps}/v_{Ph})_{ini}$ at fixed reaction conditions is often coupled to an increase in Ps_{max} and can therefore be used as an alternative screening criterion. This is in agreement with Eq. 1, which shows that an increase in β_0 and/or $1/\gamma$ (reflected by an increased rate ratio of synthetic and hydrolytic product formation) results in a higher Ps_{max} if α remains unchanged or is even decreased. A problematic feature of this screening approach is the fact that improved mutants with respect to Ps_{max} might be missed if they possess a superior α parameter, but have a similar $(v_{Ps}/v_{Ph})_{ini}$ as the wild-type enzyme. In the same way, mutants might seem improved upon screening, while actually not yielding more product if they have an unfavorably high α, which causes preference for the product over the activated acyl donor during enzyme acylation. Despite these limitations, we chose to use a screening procedure based on the described principle because of its relatively high throughput capacity, and refine the search by a second, more detailed round of analysis.

In the first round of HPLC screening, cell suspensions of mutant clones were analyzed in MTPs for the produced amount of Ps (ampicillin) and Ph (D-phenylglycine, PG) from D-phenylglycine amide (PGA) and 6-APA at a certain time. PGA was used as the acyl donor due to its high chemical stability and, consequently, low background hydrolysis under the applied reaction conditions. The peak ratio of ampicillin/PG formed after 24 h was compared to the one of a wild-type PAS2 culture. Since we could not assume all reaction mixtures to be in the initial phase of conversion, and the peak ratio changes in the course of reaction, several wild-type cultures with different induction levels were prepared in order to obtain a set of chromatograms with varying degrees of conversion. Each mutant chromatogram was then compared to the wild-type profile that showed the most similar conversion level. With this method, one 96-well MTP could be analyzed in about 5 h, leading to a throughput of maximally 4 plates per day. In order to decrease the number of clones that needs to be screened, clones expressing an active PAS2 variant were preselected

on agar plates containing D-phenylacetyl-L-leucine (PAL) as the only source of leucine. Since only 20 % of all mutants were found to be active, the screening effort for synthetic activity was reduced by a factor 5 by only analyzing mutants that could grow on PAL medium. A total of 700 of these clones was subjected to a first round of HPLC screening (Fig. 2).

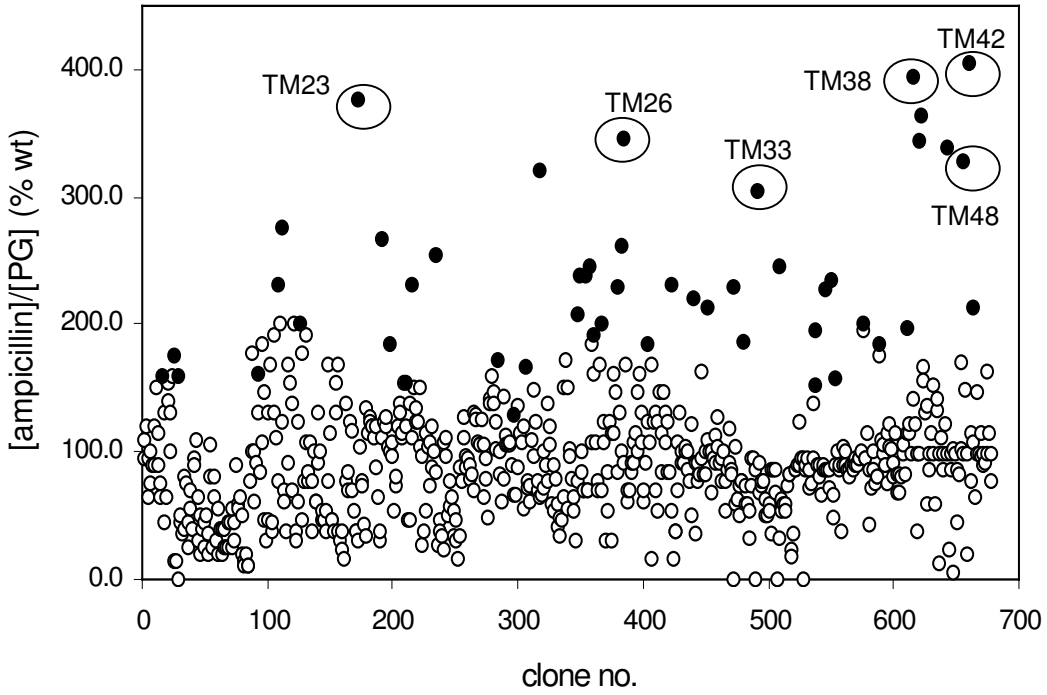


Figure 2. Screening for improved PAS2 mutants by HPLC with whole cells. Clones represented by filled dots were inspected in more detail in a second round of HPLC screening with periplasmatic extracts. Truly improved mutants that were subjected to further analysis are indicated by surrounding circles.

In preliminary experiments, we had compared the recovery of active clones by this growth selection method with the detection of activity in MTPs due to the hydrolysis of 2-nitro-5-[(phenylacetyl)amino]-benzoic acid (NIPAB), which is a well-known colorimetric substrate for penicillin acylases (Kutzbach and Rauenbusch, 1974). While all clones able to hydrolyze NIPAB were also found to grow on PAL, about 15 % of the clones utilizing PAL could not hydrolyze NIPAB, which made us decide to use PAL for the recovery of active clones. Interestingly, the 6 truly improved mutants isolated in this study (see below) could not hydrolyze NIPAB or only at extremely low rates, which would have compromised their isolation if a prescreening on NIPAB had been included.

Besides selection on PAL plates, we also isolated transformants that could grow on medium supplied with amides such as D-phenylglycine amide, α-methylphenylglycine, or D-mandelamide (racemic mixture) as the only source of nitrogen. However, none of the 500 mutants selected on these media did show promising ampicillin/PG peak ratios or high activity when tested for ampicillin synthesis. From the PAL preselection, in contrast, 50 clones exhibited enhanced synthetic behavior. The recovery of improved

mutants from PAL plates but not from nitrogen-limited media can be attributed to the fact that only traces of leucine need to be released by hydrolysis to sustain growth, while local concentrations of nitrogen are required to be in the µM-range for the formation of reasonably sized colonies. Mutants that show improved synthetic performance, but reduced hydrolytic activity may consequently not be able to grow on nitrogen-limited media.

The 50 improved mutants from the PAL preselection were tested in more detail in a second round of HPLC analysis. Here, complete progress curves for the kinetically controlled synthesis of ampicillin from PGA and 6-APA were recorded to reveal whether the respective mutant indeed allowed higher maximal product accumulation than the wild-type. Six of the mutants showed significantly improved synthetic performance and were subsequently also tested with another common β-lactam nucleophile, 7-ADCA, in the synthesis of cephalexin. Dependence on the type of acyl donor was revealed by using either PGA or the methyl ester derivative of D-phenylglycine, PGM, as the acyl moiety. Improvements were most pronounced for the synthesis of ampicillin for all mutants, particularly with PGM that allowed up to 4 times higher product accumulation than obtained with the wild-type enzyme. In general, the 6 mutants showed similar trends in the test reactions. Progress curves of triple mutant TM23 are shown as examples in Fig. 3.

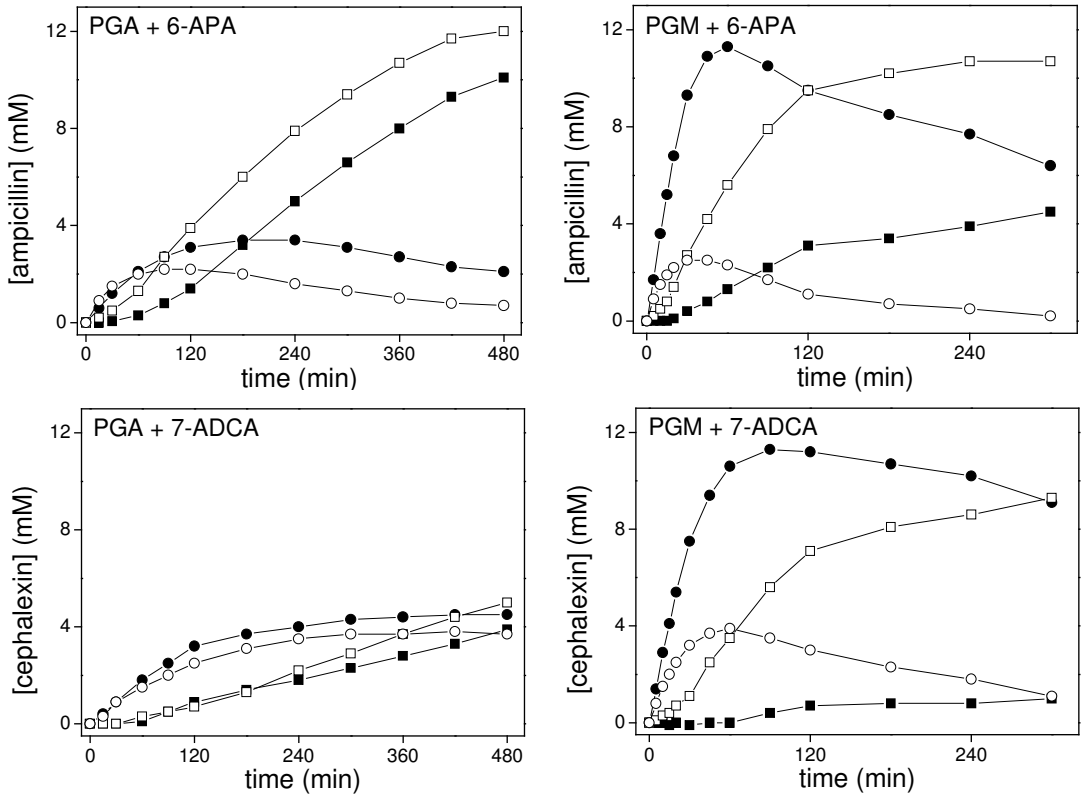


Figure 3. Synthesis of ampicillin and cephalexin by wild-type PAS2 and TM23, using 15 mM acyl donor and 25 mM β-lactam nucleophile. When PGA was used as the acyl donor, 75 nM wild-type and 7.5 µM of the mutant enzyme was used. With PGM, 250 nM wild-type and 2.5 µM mutant enzyme was added. PAS2: (○) antibiotic, (□) PG. TM23: (●) antibiotic, (■) PG.

In contrast to the correspondence in synthetic behavior, the mutants were found to differ with respect to their amino acid residue substitutions (Table 1). Remarkably, TM33 only comprised one amino acid residue change, βF24A, corresponding to the βF24A *E. coli* PA single mutant that was reported to allow high antibiotic yields (Alkema et al., 2002b). However, improvements for the *E. coli* mutant were most pronounced in the synthesis of cephalexin, while TM33 performed best in the production of ampicillin, in both cases using PGM as the acyl donor. Although TM33 did not carry mutations at the other two sites, it was confirmed as a true triple mutant by restriction analysis and the presence of altered codons for residues αR160 and αF161.

Table 1. Amino acid residue changes (bold) of improved mutants of penicillin acylase PAS2

Name	Mutations
TM23	αR160**K** αF161**L** βF24**M**
TM26	αR160**Q** αF161**L** βF24**Q**
TM33	αR160R αF161F βF24**A**
TM38	αR160**P** αF161**A** βF24**V**
TM42	αR160**Q** αF161**R** βF24**I**
TM48	αR160**P** αF161**A** βF24**Q**

The other 5 mutants contained amino acid substitutions at all three target positions. While some mutations, i.e. αR180K and αF181L, have earlier been found to influence antibiotic synthesis in *E. coli* PA single mutants (Alkema et al., 2002b; Alkema et al., 2002c), the effects of most substitutions and their combination have not been studied.

3.3. Kinetic characterization of the best mutants

While antibiotic yields were clearly improved for all mutants with PGA as the acyl donor (120 to 260 %), synthesis rates were strongly decreased, being less than 1 % of the wild-type rates. Activities with PGM were much higher for all mutants, but mutants TM26, TM42, and TM48 still maximally reached 3.6 % of the PAS2 synthesis rates. For a detailed kinetic characterization, we therefore focused on the three other mutants, TM23, TM33 and TM38 that exhibited 10 to 20 % of the wild-type activity in antibiotic formation. All experiments were carried out with purified enzyme preparations.

The steady-state parameters for the hydrolysis of the acyl donors used in this study as well as of the synthesized antibiotics are summarized in Table 2. In agreement with the low conversion rates observed in the synthesis experiments, the k_{cat} of all mutants for the tested acyl donors was decreased. TM23, TM33 and TM38 showed clear preference for hydrolysis of the ester compared to the amide substrate. Surprisingly, the wild-type enzyme was slightly more specific for the amide AD, although enzyme-catalyzed ester hydrolysis is generally expected to be faster due to the comparatively lower chemical stability of the ester bond (Polgar, 1989). Although several hypotheses

exist that, for example, relate this unexpected feature to a distortion of the normally planar amide bond towards a structure that more resembles the tetrahedral transition state that leads to the formation of the covalent acyl-enzyme intermediate (Hedstrom et al., 1992; James et al., 1980; Polgar, 1989), no structural evidence for the presence of the proposed mechanisms in PA has yet been provided.

Table 2. Steady-state kinetic parameters of penicillin acylase PAS2 and the improved mutants

Enzyme	PGA			PGM		
	k_{cat} [s^{-1}]	K_m [mM]	k_{cat}/K_m [$mM^{-1}\ s^{-1}$]	k_{cat} [s^{-1}]	K_m [mM]	k_{cat}/K_m [$mM^{-1}\ s^{-1}$]
PAS2	25	12	2.1	24	14.3	1.7
TM23	0.7	122	0.006	2.0	13.4	0.15
TM33	0.13	29	0.0053	4.8	53.4	0.09
TM38	0.15	33	0.005	1.8	13.7	0.13

Enzyme	Ampicillin			Cephalexin		
	k_{cat} [s^{-1}]	K_m [mM]	k_{cat}/K_m [$mM^{-1}\ s^{-1}$]	k_{cat} [s^{-1}]	K_m [mM]	k_{cat}/K_m [$mM^{-1}\ s^{-1}$]
PAS2	16	0.6	26.7	20	1.3	15.4
TM23	3.4	2.6	1.31	0.6	1.7	0.35
TM33	2.3	2.2	1.05	0.4	0.9	0.44
TM38	2.4	2.4	1.00	0.8	2.8	0.29

The increase in relative esterase/amidase activity of the mutants was also reflected in their α parameters for the synthesis of both ampicillin and cephalexin, which were much lower with PGM as the acyl donor than with PGA (Table 3). While with PGM, α of all mutants was smaller than the one of the wild-type enzyme, partly explaining the high antibiotic yields, α was up to 17-fold higher for the amide acyl donor. This seems to be in contradiction to the observed increase in antibiotic accumulation with PGA as the substrate. However, not only the competition between antibiotic and activated acyl donor for acylation of the enzyme determines Ps_{max}, but also the competition between the β-lactam nucleophile N and water for deacylation of the covalent acyl-enzyme intermediate (EAc), which is described by the parameters β_0 and $1/\gamma$. With 10 to 40-fold increases in β_0 and even up to 60 times higher values for $1/\gamma$ (Table 3), improvements with respect to deacylation of EAc by the tested β-lactam compounds were large enough to compensate for the relative decrease of specificity for PGA compared to the antibiotic and allowed higher yields than with the wild-type enzyme. Since formation of both reaction products, Ps and Ph, occurs via the same covalent intermediate (EAc), β_0 and $1/\gamma$ should be independent of the type of AD used, which is

in agreement with our experimental results. Whereas for PAS2, $(v_{Ps}/v_{Ph})_{ini}$ could be determined for a whole range of concentrations of the β-lactam nucleophile (1 to 200 mM), allowing the experimental determination of both β_0 and $1/\gamma$ (*Chapter 5*), only β_0 could be measured for the three mutants. This is due to the fact that the synthetic reaction was so much preferred with these enzymes that already at relatively low nucleophile concentrations the amount of hydrolytic product was below the detection limit in the initial phase of conversion. Therefore, only β_0 could be measured at 6-APA and 7-ADCA concentrations below 5 mM, whereas $1/\gamma$ needed to be obtained by fitting [Ps] *vs*. [Ph] plots that were derived from progress curves (Fig. 3) with Eq. 1, using the experimental values of α and β_0. Good agreement was observed between the data derived from curves based on the use of the amide and the ester acyl donor (Fig. 4), supporting the assumption that the formation of the acyl-enzyme intermediate is the rate-limiting step of the conversion.

Table 3. Complex kinetic parameters α, β_0 and γ of penicillin acylase PAS2 and improved mutants

Enzyme	Acyl donor	Ampicillin			Cephalexin		
		α	β_0 (mM^{-1})[a]	$1/\gamma$[b]	α	β_0 (mM^{-1})[a]	$1/\gamma$[b]
PAS2	PGA	13.2	0.5	6	7.3	0.5	59
	PGM	16.4	0.5	6	9.1	0.5	59
TM23	PGA	218	6.3	364	58.8	18.4	69
	PGM	8.8	6.3	364	2.4	18.4	69
TM33	PGA	209	10.2	286	88.8	21.5	154
	PGM	11.6	10.2	286	4.9	21.5	154
TM38	PGA	200	4.9	167	57.2	21.1	58
	PGM	7.6	4.9	167	2.2	21.1	58

[a] Mean of two independent experiments, using PGA and PGM as the acyl donor, respectively. Coefficient of variations were < 12 %.

[b] Parameter γ was obtained by fitting Ps versus Ph plots ([AD] = 15 mM, [N] = 25 mM) with Eq. 1 and the experimentally determined values for the parameters α and β_0. The given value is the average obtained from analyzing 2 different curves, one with PGA and one with PGM as the acyl donor. The coefficients of variation were < 10 % for all data. Average values are given for β_0 and $1/\gamma$, since these parameters are independent of the type of acyl donor used.

Interestingly, when 6-APA was used as the β-lactam nucleophile, improvements were more pronounced for $1/\gamma$ than for β_0, whereas with 7-ADCA the opposite situation occurred. For mutants TM33 and 38, virtually no increase in $1/\gamma$ was observed in the synthesis of cephalexin. The high values of $(v_{Ps}/v_{Ph})_{ini}$ at low concentrations of 7-ADCA indicate that the affinity for this β-lactam nucleophile became higher due to the mutations in the substrate binding pocket, while the exclusion of water from the active site at high 7-ADCA concentrations was obviously not improved. In contrast, binding of 6-APA in the active site seemed to widely suppress the nucleophilic attack of water on the acyl-enzyme intermediate.

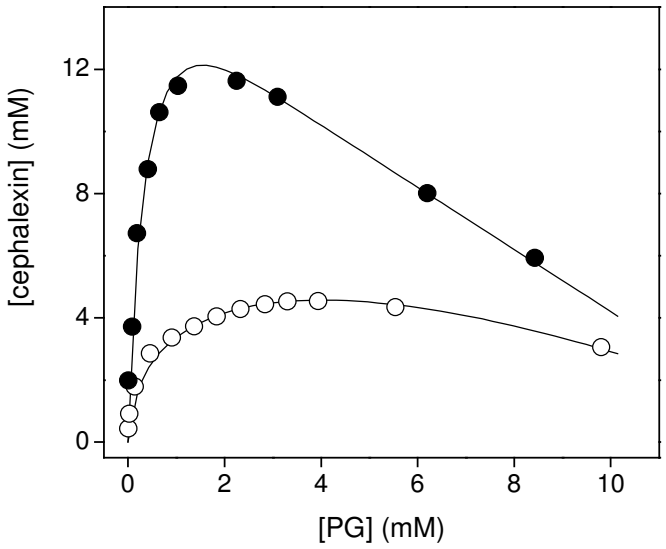


Figure 4. Formation of synthesis product (Ps = cephalexin) and hydrolytic byproduct (Ph = D-phenylglycine) during conversion of 7-ADCA and PGA (○) or PGM (●) by TM38 penicillin acylase. Lines represent fits obtained with Eq. 1, using α = 57.2 (PGA), α = 2.2 (PGM), and β_0 = 21.1 mM^{-1}. Values obtained for $1/\gamma$ were 56 (PGA) and 59 (PGM), respectively.

3.4. Antibiotic yields at high substrate concentrations

Besides by α, β_0, and γ, the maximal product yield of a given reaction is also determined by the initial substrate concentrations $[N]_0$ and $[AD]_0$. To evaluate the performance of TM23, TM33, and TM38, we therefore modeled the maximal antibiotic accumulation (Eq. 1) for a range of substrate concentrations, using the parameters given in Table 3.

From Fig. 5, it becomes apparent that all three mutants have a higher potential for the synthesis of antibiotics derived from 6-APA, particularly ampicillin, than from 7-ADCA. Although somewhat higher yields of cephalexin are predicted for low concentrations of 7-ADCA with PGA as the acyl donor, which is in agreement with our experiments (Fig. 3), wild-type PAS2 allows significantly higher yields at nucleophile concentrations above 50 mM with about 100-fold higher conversion rates. When PGM and 6-APA are used, the drawback of lower activity (10-20 % of wild-type) may though become unimportant in view of the 4 to 5 times higher ampicillin concentrations that can be reached over the whole range of substrate concentrations. This is particularly interesting since the reaction can be accelerated just by adding more enzyme, while the maximal level of product accumulation that is reached in a kinetically controlled synthesis reaction can only be increased by using a better enzyme. The unproductive loss of activated acyl donor is also drastically reduced when using the mutant enzymes: at 200 mM 6-APA concentration, about 80 % of PGM is converted to ampicillin with the mutant enzymes, while only 17 % is used for synthesis with the wild-type PAS2 penicillin acylase, the remainder being hydrolyzed to D-phenylglycine (Fig. 5).

In conclusion, our study demonstrates that focusing random mutagenesis to amino acid residues that are expected to affect the catalytic behavior of the target enzyme based on sequence and structural information is an efficient tool for rising the frequency of biocatalysts with a desired property in a mutant library. With this method, three mutant penicillin acylases carrying different amino acid substitutions were recovered that allow significantly higher ampicillin yields with PGM than the wild-type enzyme, while retaining relatively high catalytic activity.

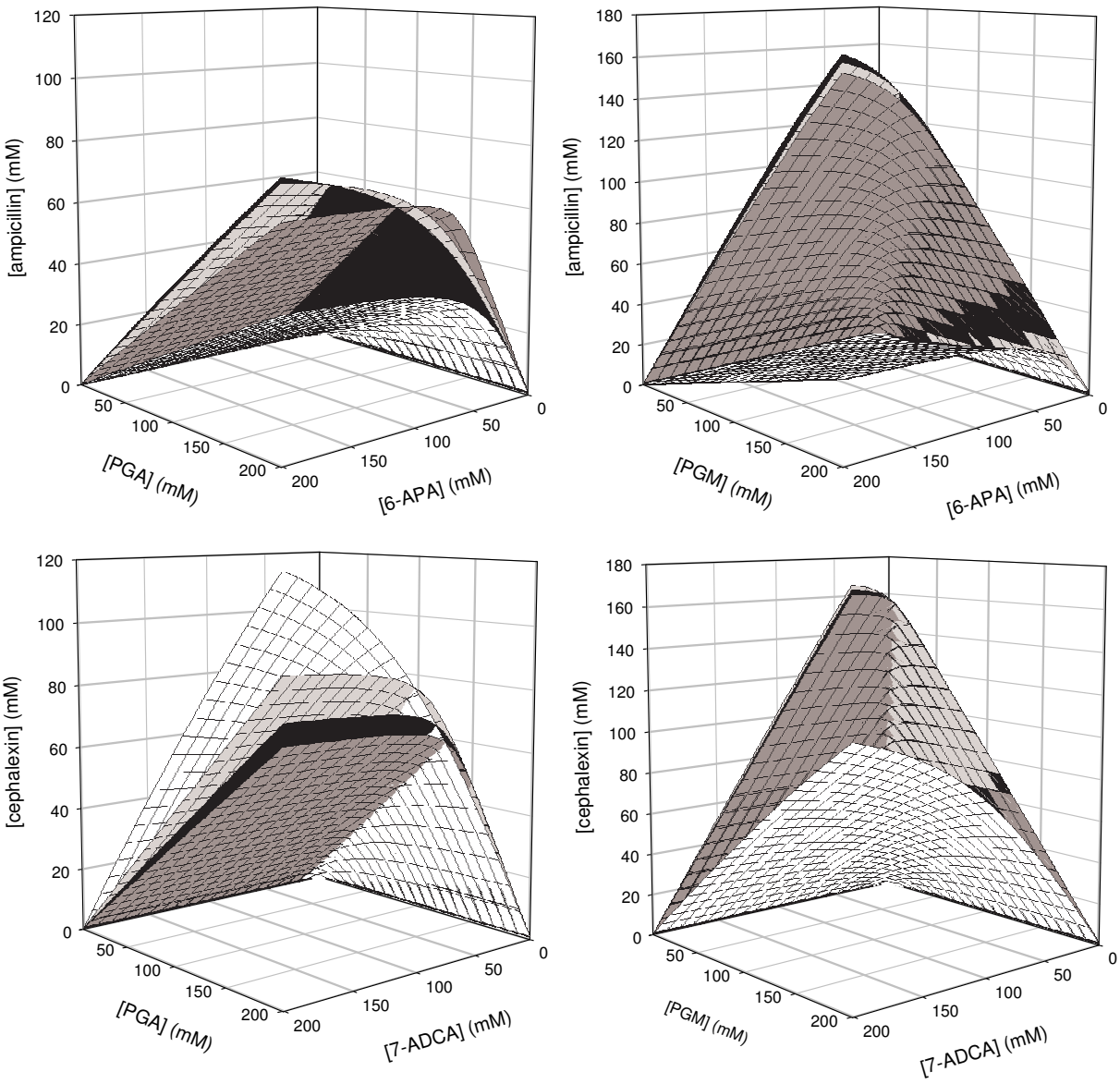


Figure 5. Maximal accumulation of ampicillin and cephalexin in PA-catalyzed, kinetically controlled conversions, using different substrate concentrations. Data were simulated with Eq. 1, using the complex kinetic parameters α, β_0 and γ given in Table 3. ☐ PAS2; ■ TM23; ■ TM33; and ■ TM38.

7

Mining the metagenome for glucovanillin glucosidases

We explored a method to discover new glucovanillin glucosidases for the biocatalytic production of natural vanilla flavor. For this purpose, gene libraries were constructed from the DNA of a microbial community that colonized the surface of green beans of *Vanilla planifolia*, and a loam soil metagenome. The libraries were searched for clones exhibiting the desired enzymatic activity with a growth selection method based on the direct use of glucovanillin as the selective carbon source. With this strategy, nine different glucovanillin-metabolizing clones were isolated. Sequence analysis suggested that three of these clones encode putative glucosidases, two of them not showing homology to any known glycosidase and, therefore, being likely to belong to novel classes of glycosidases. The concomitant recovery of relatively large numbers of false-positive clones with a stable phenotype indicates that the applied selection method was not completely selective for the target activity. Possible causes and solutions for this problem are discussed.

1. INTRODUCTION

Vanillin is one of the most important flavor compounds in the food and beverages industries as well as in the cosmetic and pharmaceutical sector. Traditionally, vanillin is extracted from processed beans of vanilla, a tropical orchid plant that originates from Mexico and is nowadays mainly cultivated in Madagascar, Indonesia, and Comoros (Priefert et al., 2001). Cultivation of this plant outside of Mexico is laborious, since it involves manual pollination of flowers because of the absence of a natural pollinator. But also the harvest and the subsequent cumbersome processing of the mature green beans, the so-called curing, complicates the production of the aroma compound as well as the low amount of vanillin that is eventually present in the cured, dark brown pods [2 % (w/w) of the dry matter; Vandamme and Soetaert, 2002]. Therefore, natural vanillin is one of the most expensive flavors (> 4,000 € kg^{-1}) with an annual production of only 20 tons. In contrast, about 20,000 tons of synthetic (nature-identical) vanillin are chemically prepared per year, e.g. from lignin, a waste product of the pulp and paper industries, at a price as low as 12 € kg^{-1}. This considerable difference in price, the better taste of natural vanillin, and the increasing trend towards the use of non-synthetic products (Armstrong and Yamazaki, 1986) explains the economic interest in the production of natural vanillin.

Besides using the traditional processes, European and US food legislation has opened an attractive way for obtaining natural flavors by defining a flavor to be "natural" as long as it is produced by "appropriate physical processes or enzymatic or microbiological processes from material of vegetal or animal origin" (EU Flavor Directive 88/388/EEC). This means that procedures may be devised that use precursor molecules easily available from natural products, which are then converted to the desired flavor by a suitable enzyme, (micro)organism or cell culture, even though the starting materials might not be the traditional source of the respective flavor. Several methods for the production of natural vanillin have been developed on account of these regulations, using natural and relatively cheap precursors such as ferulic acid (present

in sugar beet pulp and cell walls of woods and grasses), eugenol (a main constituent of clove oil) or phenolic stilbenes (in spruce bark) (Fig. 1). Also *de novo* biosynthesis in (genetically modified) microbial or *Vanilla* plant tissue cultures is under investigation [for a recent review on the biotechnological production of vanillin see Priefert et al. (2001)]. However, most of the mentioned processes are not (yet) economical and additionally suffer from the fact that their products lack the full aroma of natural vanilla (Krings and Berger, 1998).

Especially in the latter respect, another approach seems to be promising, constituting a biotechnological upgrade of the traditional vanilla curing process. In vanilla pods, the major precursor of vanillin is its β-D-glucoside (glucovanillin) from which it is released by the action of β-glucosidase(s) (Arana, 1943; Fig. 1). In traditional vanilla production, flavor development is based on slow thermal and naturally induced enzymatic reactions that can take 2-6 months (Ramachandra Rao and Ravishankar, 2000; Ranadive, 1994). By directly extracting the flavor compound from the ripe but uncured green beans in the presence of commercially available enzyme preparations with β-glucosidase activity, the lengthy curing process can be reduced from several months to a few hours. Due to the exogenous enzyme activity, glucovanillin can be hydrolyzed faster and more completely than during traditional curing, leading to yield improvements of up to 24 % (Ranadive, 1992; Ruiz-Terán et al., 2001; Brunerie, 1998). Besides vanillin, other important flavor compounds such as *p*-hydroxybenzoic acid, *p*-hydroxybenzaldehyde, *p*-hydroxybenzyl alcohol, vanillic acid, and vanillyl alcohol are released from their glucoside precursors and contribute to the characteristic natural vanilla aroma. A set of β-glucosidases with different substrate ranges concerning the aglyconic part of β-D-glucosides could therefore constitute a useful means to tune the overall flavor obtained from biotechnologically cured vanilla pods. While the origin of β-glucosidase activity in vanilla beans has not been thoroughly established yet, Röling et al. (2001) isolated a variety of strains of the genus *Bacillus* from the surface of cured vanilla pods that exhibit cell-bound β-glucosidase activity. This suggests that besides plant enzymes, bacterial enzymes may be involved in hydrolysis of flavor glucosides, which makes the bacterial community attached to vanilla pods a promising source of enzymes for biotechnological curing.

In this study, we explored this idea by using genomic DNA isolated from enrichment cultures that were inoculated with vanilla pod microorganisms to construct two different expression gene banks. These gene banks as well as a library previously prepared from DNA isolated from loam soil (*Chapter 2*) were screened for glucovanillin glucosidase activity by a growth–based assay. Selection methods generally allow the highest throughput in assaying gene banks. Hundred thousands of clones can be analyzed on one single agar plate as only a few recombinants – the positives – will be able to withstand the selective pressure and form colonies. Selection media need to be designed carefully in order to prevent non-specific background growth and the formation of false-positive colonies due to undesired metabolic activities. Here, we report on the development of a growth-based assay for glucovanillin glucosidase. Although the procedure still has its pitfalls, our strategy led to the isolation of 3 clones with (putative) β-glucosidase activity, carrying enzyme-

coding sequences that only show low homology to known enzymes or are similar to hypothetical proteins of unknown function.

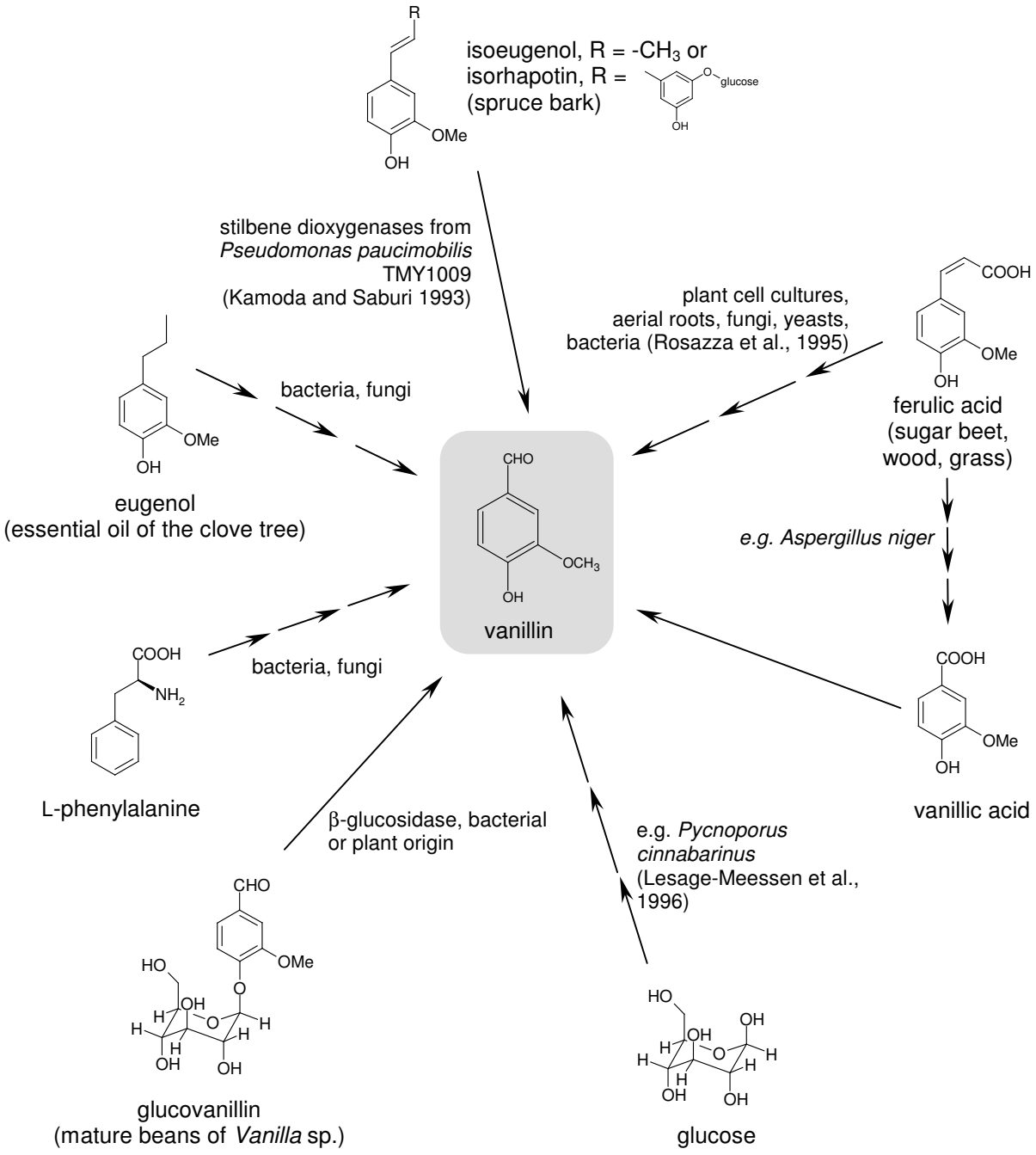


Figure 1. Biotechnological production of natural vanillin. Besides from its traditional source, i.e. mature vanilla beans, vanillin is obtained from a number of cheap precursor molecules by biotransformations, using different microorganisms or plant cell cultures. According to the EU Flavor Directive 88/388/EEC, vanilla obtained in this way may still be labeled *natural*.

2. MATERIALS AND METHODS

2.1. Environmental samples

Green (non-cured) pods of *Vanilla planifolia* were a gift from Quest International (The Netherlands). The pods originated from the Indonesian vanilla curing company Djasula Wangi and were used as a source of microorganisms that colonize the surface of this plant material. After harvest in April 2000, vanilla pods were shipped at ambient temperature to The Netherlands before they were stored at –20°C in the laboratory. Loam soil was collected from an agricultural field as described in *Chapter 2*.

2.2. Enrichment and environmental gene bank construction

To release the attached microorganisms, a protocol from Röling et al. (2001) was used that is optimized for the extraction of microorganisms from vanilla beans. About 100 g of vanilla pod material was chopped into 2-cm pieces and incubated in 200 ml of 0.85 % NaCl for 6 h with vigorous shaking at room temperature in the presence of 20 g of glass beads (2 mm diameter). The number of recovered microorganisms was determined with serial dilutions of a small aliquot of the cell suspension, which were plated on 3 % tryptic soy agar (TSA, Difco). Plates were incubated at 30°C for 3-5 days before colonies were counted.

The remaining suspension was centrifuged at 5,000 g for 15 min to collect the released organisms that were subsequently washed twice and resuspended in 10 ml of 0.85 % NaCl. From this cell suspension, 50 µl were used to inoculate 1 ml of rich enrichment medium (3 % liquid TSA medium) and 1 ml of selective enrichment medium [SEM: minimal medium (*Chapter 2*) containing 20 mg l^{-1} yeast extract (BBL) and 2 mM β-D-glucovanillin as the sole carbon source]. After 1 day (TSA culture) or 3 days (SEM culture) of incubation at 30°C and orbital shaking at 150 rpm, the complete precultures were transferred to 100 ml of fresh medium. No growth could be observed in a control SEM culture to which no glucovanillin was added, indicating that the turbidity of the real SEM culture was not caused by background growth on the supplied yeast extract. After incubation of the 100-ml cultures for another 1 or 3 days, respectively, cells were collected by centrifugation (5,000 g, 15 min) and the pellets were used for DNA extraction and gene bank construction in *E. coli* TOP10 as described for enrichment cultures in *Chapter 4*. The high-copy number plasmid pZero-2 (Invitrogen) that confers kanamycin resistance was used as the cloning vector. Transformants were spread on LB (+ 10 mg l^{-1} kanamycin) medium and colony-forming units (cfu) were enumerated after overnight incubation at 30°C. Gene banks (insert sizes: 2 to 7 kb) were stored at –20°C as amplified plasmid preparations (High Pure Plasmid Isolation kit, Roche).

The diversity of bacterial genomes comprised in the two genomic DNA preparations used for library construction was estimated by denaturing gradient gel electrophoresis (DGGE) according to *Chapter 2*.

Depletion of glucovanillin in the different enrichment cultures was monitored by high-performance liquid chromatography (HPLC). HPLC analyses were carried out

using a 10-cm Chrompack C18 column (5 mm diameter) in connection with Jasco PU-980 pumps and a Jasco MD-910 detector set at 214 nm. Compounds were isocratically eluted at a flow rate of 1 ml min^{-1} with a solution of 340 mg l^{-1} sodium dodecylsulfate and 680 mg l^{-1} $KH_2PO_4 \cdot 3H_2O$ in a 30:70 (v/v) acetonitrile/water mixture of pH 3.0 (adjusted with diluted phosphoric acid). Under these conditions, the peaks of glucovanillin and its hydrolysis product vanillin were clearly separated, with retention times of 1.2 and 1.9, respectively.

2.3. Selection of β-D-glucovanillin metabolizing clones

Transformants were washed from plates with minimal medium, washed twice, and plated on a medium selective for recombinants that are able to utilize β-D-glucovanillin as a growth substrate. The solid medium consisted of minimal medium supplied with 15 g l^{-1} agar, 5 mg ml^{-1} leucine to complement the leucine auxotrophy of the used *E. coli* host strain, 10 mg l^{-1} kanamycin, and 2 mM β-D-glucovanillin as the only source of carbon. Besides the two gene banks that were prepared from DNA isolated from enrichment cultures, a metagenomic loam soil library (*Chapter 2*) was also subjected to growth selection. The latter gene bank comprised 80,000 clones with an average insert size of 5.2 kb. As amplified gene banks were used for selection, plasmids of positive clones were isolated and subjected to enzymatic restriction analysis in order to identify unique clones. Plasmid-encoded activity of positives was subsequently verified by retransforming the respective plasmids to *E. coli* TOP10 and plating the recombinants on selection medium again.

2.4. Substrate profiling

The different β-D-glucovanillin degrading clones isolated in this study were simultaneously tested with various substrates in a 96-well microtiter plate (MTP) format. Cell material from single colonies was used to inoculate 250 µl of LB plus 10 mg l^{-1} kanamycin. After overnight incubation at 30°C with orbital shaking at 200 rpm, 40 µl samples of each well were transferred to fresh MTPs and mixed with 10 µl of 10 mM substrate solutions. The conversion of substrates was followed in time with a Powerwave X plate reader (Bio-Tek Instruments). The release of *p*-nitrophenol ($\Delta\varepsilon$ = 10.077 mM^{-1} cm^{-1}) from *p*-nitrophenyl-β-D-glucopyranoside, *p*-nitrophenyl-α-D-glucopyranoside, and *p*-nitrophenyl-β-D-galactopyranoside, and of *o*-nitrophenol from *o*-nitrophenyl-β-D-glucopyranoside ($\Delta\varepsilon$ = 2.295 mM^{-1} cm^{-1}) was monitored at 405 nm, whereas the formation of vanillin ($\Delta\varepsilon$ = 11.201 mM^{-1} cm^{-1}) due to the hydrolysis of glucovanillin was recorded at 350 nm. Almond β-glucosidase (Sigma) was used as a positive control.

2.5. Sequence analysis

Insert sequences of the recombinant plasmids were determined at BaseClear Holding BV (Leiden, The Netherlands). Open reading frames (ORFs) were identified with the ORF Finder service available at the NCBI web page (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). For translation to protein sequences, the Bacterial Code was selected, allowing ATG, GTG, TTG, ATT, and CTG as alternative

start codons. The minimal ORF length was set to 300 bp. When alternative start sites for the same gene were encountered, the largest ORF was selected. ORFs were translated and used as queries in BLAST searches of the DDBJ/EMBL/GenBank database entries. The N-terminus (70 amino acid residues) of each derived protein sequence was furthermore analyzed for the presence of a cleavable N-terminal signal peptide or membrane anchor with the programs SignalP (Henrik et al., 1997) and PSORT (Nakai and Horton, 1999).

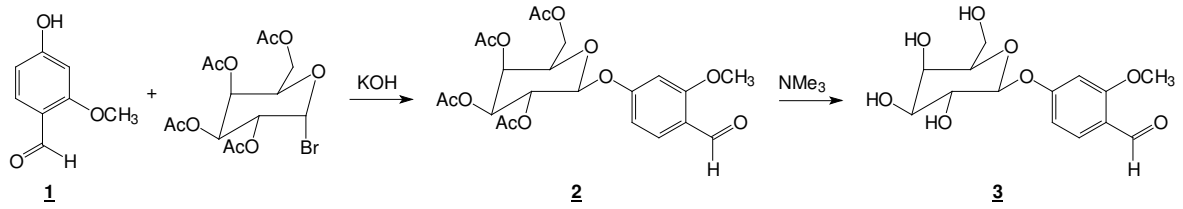


Figure 2. Synthesis of glucovanillin. Vanillin (1) and 2,3,4,6-tetra-*O*-acetyl-α-D-glucopyranosyl bromide are reacted to tetra-*O*-acetyl-glucovanillin (2) that yields glucovanillin (3) upon deprotection.

2.6. Chemicals

Colorimetric substrates were purchased from Sigma. Glucovanillin was prepared by the following procedure. Vanillin (Fig. 2, compound 1, 15.2 g, Sigma) was dissolved in 100 ml dioxane. 65 ml 1 M KOH was added to give a yellow solution of pH 9.7. To this, a solution of 2,3,4,6-tetra-*O*-acetyl-α-D-glucopyranosyl bromide (40 g, Sigma) in 100 ml acetone was added dropwise. During this process, the solution was sparged with N_2-gas and the pH was kept above 9.5 by adding 1 M KOH (60 ml in total). The mixture was stirred overnight. The resulting brown mixture (pH 5.9) was extracted with 4 100-ml portions of $CHCl_3$. Evaporation of the $CHCl_3$ gave a brown oil that was taken up in 200 ml MeOH. Tetra-*O*-acetyl-glucovanillin (compound 2, Fig. 2) precipitated as a pure white product. After recrystallization from MeOH, filtration, rinsing with Et_2O and drying, 14.25 g of compound 2 were recovered. 300 MHz ^{1}H NMR (DMSO*d6*) α: 1.984, 2.002, 2.008 (s, 12H, CH_3); 3.821 (s, 3H, OCH_3); 4.062-4.275 (m, 3H, CH*sugar*); 4.985-5.129 (m, 2H, CH*sugar*); 5.403 (m, 1H, CH*sugar*); 5.613 (d, 1H, CH*sugar*); 7.290 (d, 1H, CH*phenyl*); 7.476 (s, 1H, CH*phenyl*); 7.574 (d, 1H, CH*phenyl*); 9.886 (s, 1H, CHO). 300 MHz ^{13}C NMR (DMSO*d6*) α: 17.679, 17.776, 17.857, 17.860 (CH_3); 65.324, 67.946, 68,431, 69.208 (CH*sugar*); 58.964 (CH_2*sugar*); 94.956 (C*anomer*); 108.793, 113.843, 122.064 (CH*phenyl*); 129.347, 147.262, 148.006 (C*phenyl*); 166.343, 166.699, 166.990, 167.362 (CO); 189.048 (CHO).

Compound 2 (14.0 g) was suspended in 200 ml MeOH, and 40 ml of a 45 % solution of NMe_3 in H_2O was added. This mixture was left standing overnight at room temperature. The resulting white mass was filtered off and dried under vacuum, yielding pure glucovanillin (Fig. 2, compound 3, 6.5 g). 300 MHz ^{1}H NMR (DMSO*d6*) α: 3.110-3.700 (m, 5H, CH*sugar*); 3.825 (s, 3H, CH_3); 4.575 (m, 1H, CH*sugar*); 5.050-5.175 (m, 4H, OH); 5.375 (m, 1H, CH*anomer*); 7.250 (d, 1H, CH*phenyl*); 7.422

(s, 1H, CH*phenyl*); 7.502 (d, 1H, CH*phenyl*); 9.850 (s, 1H, CHO). The motherliquor yielded starting compound 2 (25 %).

3. RESULTS AND DISCUSSION

3.1. Construction of enrichment gene banks

On arrival in our laboratory, green vanilla pods from Djasula Wangi were found to harbor only about 15,000 detachable cfu per bean. We think that this low number is due to the lengthy transport and storage of beans at –20°C prior to analysis, which may have caused part of the microbial population to die or to turn into a non-cultivable state. Of course, a larger number of organisms than detected might in fact have been present, including bacteria that cannot grow under the applied growth conditions. However, Röling et al. (2001) who studied the same type of green beans mostly detected bacteria on the bean surface that can be grown on TSA, which supports the assumption that only few bacteria were indeed present. Consequently, more than 1 million of beans was needed to yield enough microorganisms for DNA extraction and gene bank construction (10-100 µg genomic DNA required), assuming an average DNA content of 1 fg per cell. In view of this impossibly large number, we decided to carry out an enrichment step prior to DNA isolation. In one case, liquid TSA medium was used for cultivation to allow the growth of a large variety of cells. For the selection of organisms actually exhibiting β-D-glucosidase activity towards our target substrate, a second culture was inoculated, using a selective enrichment medium (SEM) with glucovanillin as the only source of carbon. In order to maintain a high degree of diversity also in this case, cultivation times were kept short with only one transfer to fresh medium in-between. Under the applied conditions, no chemical hydrolysis of glucovanillin could be detected by HPLC analysis.

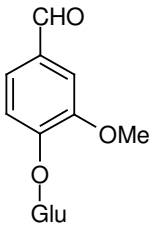
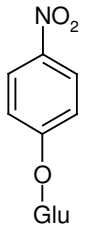
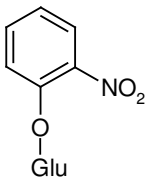
Diversity was estimated by plating serial dilutions of the two enrichments on TSA agar medium and by DGGE analysis of the recovered genomic DNA. With both techniques, about 9 abundant types of organisms were detected in the TSA enrichment, while only 4 dominant species were found after cultivation in selective medium. Cells recovered from the TSA culture yielded a total of about 250 µg of genomic DNA as estimated from agarose gels, whereas only 50 µg could be isolated from the SEM culture due to lower cell densities. From small aliquots of the isolated DNA, gene banks comprising 45,000 (TSA enrichment) and 80,000 clones (SEM enrichment) were prepared with insert sizes of 2-7 kb.

3.2. Selection of glucovanillin-metabolizing clones + substrate profiling

Three factors were considered when optimizing the solid selection medium for gene bank testing: the concentration of glucovanillin, which was added as the sole source of carbon, needed to sustain the formation of reasonably sized bacterial colonies (1), the chemical stability of glucovanillin under the applied experimental conditions (2), and the potential toxicity of vanillin released upon hydrolysis of glucovanillin, since vanillin is a well-known antimicrobial agent (Cerrutti and Alzamora, 1997) (3).

The optimal concentration of glucovanillin was approximated by the requirements of *E. coli* TOP10 cells for glucose under the applied growth conditions, which was experimentally determined to be 1 mM. In order to compensate for a possibly lower affinity towards the selective substrate, glucovanillin was added at a final concentration of 2 mM.

Table 1. Relative activities of the 9 glucovanillin-utilizing clones on different β-glucosidic substrates

Construct	Relative activity (%)[a] on		
			
	glucovanillin	*p*-nitrophenyl-β-D-glucopyranoside	*o*-nitrophenyl-β-D-glucopyranoside
pTSA1	52	46	100
pTSA3	100	88	87
pTSA4	67	100	83
pSEM1	53	100	89
pSEM2	62	67	100
pSEM3	51	71	100
pLOAM1	83	54	100
pLOAM2	34	56	100
pLOAM3	51	63	100
Almond β-glucosidase	91	54	100

[a] Activity is given with respect to the substrate that was converted the fastest (7.4 mU, pTSA1; 13.4 mU, pTSA3; 10.1 mU, pTSA4; 12.1 mU, pSEM1; 4.5 mU, pSEM2; 19.6 mU, pSEM3; 9.2 mU, pLOAM1; 3.2 mU, pLOAM2; 5.7 mU, pLOAM3) with 1 mU defined as the conversion of 1 nmol substrate min^{-1} in 1 ml of cell suspension of OD_{600}. For each clone, one master culture was prepared, of which aliquots were used for the different test reactions. Almond β-glucosidase converted 3.1 µmol substrate min^{-1} mg^{-1}.

No chemical hydrolysis of glucovanillin could be detected by HPLC analysis during 2 weeks of incubation at 30°C. When growing *E. coli* TOP10 on minimal agar medium supplied with 5 mg ml^{-1} leucine, 1 mM glucose and vanillin concentrations of up to 3 mM, no inhibitory effect of the aromatic compound on the growth of *E. coli* TOP10 could be detected.

The optimized selective medium was used for the selection of glucovanillin-metabolizing clones comprised in the two enrichment gene banks as well as in a gene bank constructed from metagenomic soil DNA (*Chapter 2*), leading to the isolation of a total of 11 unique positive clones as established by restriction analysis. After retransformation of their isolated plasmids, two of the constructs, pTSA2 and pSEM4, did not confer the ability to grow on glucovanillin to their host cells and were, therefore, not studied any further. The other 9 plasmids resulted in clones with a stable glucovanillin degradation phenotype that formed colonies after 2 to 3 days of incubation on the selection medium.

To rapidly test the functional diversity of the different clones, they were grown in replicates in 96-well plates and incubated with a number of substrates yielding a product upon conversion that can be detected spectrophotometrically (Table 1). All clones were highly specific for β-glucosidic bonds and could not convert the α-derivative of *p*-nitrophenyl-D-glucopyranoside. They were also specific for the sugar part (glucose) of the test substrates, except one clone, pSEM2, that was also active on *p*-nitrophenyl-β-D-galactopyranoside. Furthermore, this clone was the only one with a broader substrate spectrum concerning the aglyconic part of the supplied substrates, accepting also the bulkier indole group of the chromogenic substrates 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) and the respective glucopyranoside X-Glc. Concerning the substrates listed in Table 1, some variation in relative activities could be detected between different clones, although differences were rather small.

3.3. Sequence analysis

Inserts of the 9 constructs were sequenced and ranged between 1.6 and 5 kb in size (Table 2). Unexpectedly, only 2 insert sequences contained open reading frames (ORFs) with homology to known glycosidases or other hydrolyzing enzymes (pTSA4 and pLOAM1). *E. coli* (pTSA4) though carried only a partial β-glucosidase-coding sequence, suggesting that the ability to metabolize glucovanillin was caused by another ORF. The absence of putative coding sequences that exhibit homologies that could readily explain the observed activity prompted us to focus on the details of β-glucoside metabolism in *E. coli*.

While β-galactosides are easily transported into the cytoplasm of *E. coli* cells via a permease (LacY), K-12-derived *E. coli* strains lack a transporter for β-glucosides, compromising the uptake of these compounds. These strains still contain a cryptic operon (*bgl* operon) that encodes a specific enzyme II-enzyme III complex (*bglS*) for the transport of β-glucosides into the cytoplasm via the PTS (Fig. 3). Normally, the *bgl* operon is not expressed due to a mutation in its promoter region. If expression can though occur under certain conditions, the *bglS* gene product leads to the uptake and phosphorylation of β-glucosidic substrates. This can be due to a reverting mutation in the promoter sequence or due to a heterologous transcription regulator that directs the *E. coli* RNA polymerase to the mutated *bgl* promoter and causes transcription. The phosphorylated compounds can then either be hydrolyzed by another enzyme of the *bgl* pathway, the phospho-β-glucosidase BglB, or by the constitutively expressed phospho-β-glucosidase BglA. Complementation of an *E. coli* K-12 host strain's

inability to metabolize β-glucosides can therefore occur in four ways: (1) expression of a transcription regulator that allows expression of a transporter from the defective *bgl* promoter, (2) expression of a transporter molecule that can transport and phosphorylate the substrate, (3) simultaneous expression of a non-phosphorylating transporter and a β-glucosidase, and (4) production of a secreted β-glucosidase. In case 1 and 2, active clones would in fact be false-positives in a screening for β-glucosidases, while case 3 and 4 would yield the desired class of enzymes. In principle, Bgl^+ revertants can also occur spontaneously, which we think is the reason for the two false-positive clones that were isolated after the first round of selection, but lost activity when isolated constructs were retransformed to fresh *E. coli* cells.

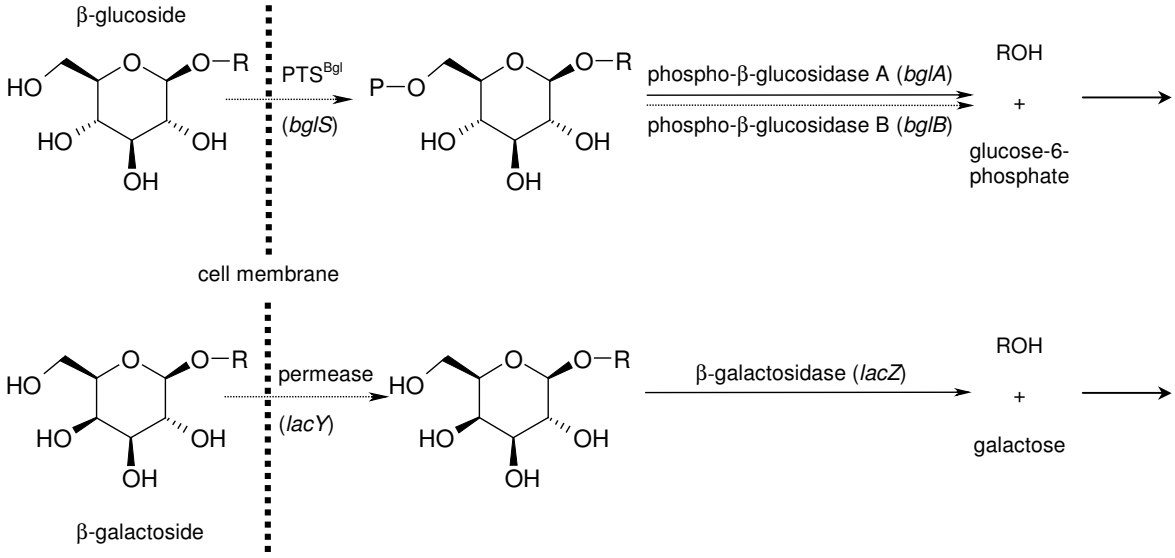


Figure 3. Uptake and hydrolysis of β-glycosides in *E. coli*. While β-glucosides are phosphorylated upon transport into the cytoplasm, β-galactosides are internalized without modification.

It is difficult to judge from its structure whether our selective substrate, glucovanillin, can (actively or passively) be taken up into *E. coli* TOP10 cells. On one hand, glucovanillin is very similar to salicin (2-[hydroxymethyl]phenyl) and arbutin (4-hydroxyphenyl-β-D-glucopyranoside), which are both known to rely on transport into the cell via the PTS^{Bgl} system. On the other hand, *E. coli* can readily internalize the β-glucoside X-Glu (Lebbink et al., 2000) even without expression of the specific PTS components, possibly due to the relaxed specificity of the LacY permease. In view of the properties of the obtained clones and their insert sequences, it can though be anticipated that glucovanillin cannot pass the cell membrane of wild-type *E. coli TOP* 10 cells, since only a single construct (pLOAM1) with homology to a - probably secreted - glycosidase was recovered (ORF1). Besides ORF1, the insert of pLOAM1 contains two other ORFs that exhibit homology to glycosidases, namely to an endoglucanase and a glycosyltransferase, which might also be responsible for the

observed activity on glucovanillin. Most of the other heterologous DNA fragments that were recovered, however, encoded putative transporters (pTSA1), transcription regulators (pSEM1, pLOAM2, and pLOAM3), or both (pTSA4), which suggests that they are false-positives in the sense of case 1 and 2, respectively, as described above. This hypothesis is supported by the observation that crude cell extracts of these clones were inactive in glucovanillin hydrolysis, indicating that an intact membrane is a prerequisite for substrate conversion. Consequently, the expression of a phosphorylating transporter either from the *bgl* operon or from a heterologous coding sequence in combination with the presence of BglA or BglB seems to enable the 5 clones to metabolize β-glucosides.

Table 2. Sequence analysis of recombinant plasmids conferring the ability to grow on glucovanillin to *E. coli* TOP10. Plasmid names of clones with possible β-glucosidase activity are given in bold.

Plasmid (insert size, bp)	ORF (no. of encoded amino acids)[a]	Closest database homologue (accession number)	Organism	Identity[b]	Signal[c]
pTSA1 (2,372)	ORF1 (167)	Putative esterase/lipase (NP_626380)	*Streptomyces coelicolor*	47%, 116/267*	No
	ORF2 (271)	**Lipoprotein, periplasmatic ABC transporter component (NP_693019)**	*Oceanobacillus iheyensis*	51%, 269/274	Yes
	ORF3 (288)	Hypothetical protein (ZP_00125017)	*Pseudomonas syringae*	41%, 236/452**	No
pTSA3 (1,945)	ORF1 (281)	Putative dehydrogenase	*Deinococcus radiodurans*	63%, 250/347*	No
	ORF2 (298)	**Multicopper oxidase (NP_768850)**	*Bradyrhizobium japonicum*	41%, 283/445	No
	ORF3 (116)	*ccdB* (vector) fusion, besides no homology			No
pTSA4 (5,015)	**ORF1** (337)	**Sugar ABC transporter (NP_244315)**	*Bacillus halodurans*	61%, 336/335	Yes
	ORF2 (194)	Transporter (NP_832140)	*Bacillus cereus*	30%, 195/194	Yes
	ORF3 (201)	**Transcription regulator (NP_762528)**	*Vibrio vulnificus*	27%, 212/276	No
	ORF4 (174)	Hypothetical protein (S32217)	*Bacillus megaterium*	50%, 163/216	No
	ORF5 (105)	β-Glucosidase (NP_348033.1)	*Clostridium acetobutylicum*	80%, 105/473**	No
	ORF6 (196)	*ccdB* (vector) fusion, besides no homology			No
pSEM1 (2,081)	**ORF1** (308)	**Transcription regulator, arabinose operon control protein (NP_655023)**	*Bacillus anthracis*	31%, 259/300	No

Table 2 – continued

Plasmid (insert size, bp)	ORF (no. of encoded amino acids)[a]	Closest database homologue (accession number)	Organism	Identity[b]	Signal[c]
pSEM2 (1,634)	**ORF1** (181)	**Hypothetical protein (NP_388796)**	*Bacillus subtilis*	42%, 175/189	Yes
pSEM3 (2,111)	ORF1 (262)	Putative ion transporter (NP_391010)	*Bacillus subtilis*	45%, 253/375	Yes
	ORF2 (142)	**Hypothetical protein (NP_834570)**	*Bacillus cereus*	54%, 130/137	No
	ORF3 (155)	Glucose-6-phosphate isomerase (NP_391013)	*Bacillus subtilis*	83%, 148/458**	No
	ORF4 (275)	*ccdB* (vector) fusion, besides no homology			No
pLOAM1 (2,261)	**ORF1** (130)	**Probable secreted glycosyl hydrolase (NP_868755)**	*Pirellula* sp.	36%, 66/446	No
	ORF2 (386)	**Endoglucanase E1 precursor (P54583)**	*Acidothermus cellulolyticus*	25%, 75/562	No
	ORF3 (174)	**Glycosyltransferase (ZP_00006152)**	*Rhodobacter sphaeroides*	30%, 174/657	No
	ORF4 (107)	No significant homology			No
pLOAM2 (3,235)	**ORF1** (165)	**Transcription regulator (NP_268160)**	*Lactococcus lactis*	26%, 120/222	No
	ORF2 (410)	No significant homology			Yes
	ORF3 (824)	SecA protein (NP_840882.1)	*Nitrosomonas europaea*	70%, 829/909	No
	ORF 4 (260)	No significant homology			No
	ORF5 (262)	No significant homology			No
pLOAM3[d]	ORF1 (154)	Putative regulator (AE005668_3)	*Escherichia coli*	98%, 154/241**	No
	ORF2 (282)	**Transcriptional regulator (AE005668_4)**	*Escherichia coli*	99%, 225/225	No

[a] ORFs with alternative start sites for the same gene were counted as one and are represented by the largest ORF. ORFs thought to be responsible for the observed activity are given in bold.

[b] The numbers behind the comma indicate the length of the homologous amino acid sequence stretch with respect to the protein length of the database hit. ORFs marked by asterisks encode presumably truncated proteins that are located at one of the ends of their inserts with a few amino acid residues encoded by the vector molecule: truncated C-terminus (*), truncated N-terminus (**).

[c] Derived protein sequences that were found to carry a signal sequence by both PSORT (Nakai and Horton, 1999) and SignalP (Henrik *et al*., 1997). Signal sequences may be cleaved off during export to the periplasm (Gram-negative bacteria) or secretion, may serve as membrane anchors, or may be a characteristic feature of lipoproteins.

[d] Insert could only be partially sequenced starting from its 5' end due to the presence of an extremely GC-rich region at its 3' end.

At first sight, the activity of *E. coli* (pTSA3) on glucovanillin is not easily explained with its two complete ORFs encoding a dehydrogenase and an oxidase homologue. In a protein BLAST search, ORF1 aligned to 85.7 % of a sequence motif (COG1250) present in 3-hydroxyacyl-CoA dehydrogenases that are typically involved in lipid transport and metabolism. ORF2, on the other hand, showed clear similarity to laccases and other ligninolytic phenol oxidases present in *Basidomycete* sp. (fungi). Hypothetically, the concerted action of both gene products might thus lead to the observed conversion of β-glucosides: by interfering with the *E. coli* lipid metabolism, the ORF1 gene product may lead to the permeabilization of the cell membrane, facilitating the uptake of e.g. glucovanillin, while the ORF2-encoded enzyme might cause the further degradation (via oxidation?) of the substrate. Since laccases are usually secreted enzymes (Bollag et al., 1988), substrate conversion may also be carried out in the extracellular space, although no signal sequence was detected in ORF2.

Besides pLOAM1, two other constructs may be true positives: pSEM2 and pSEM3. Both insert sequences carry an ORF with clear homology to rather small (< 200 aa) hypothetical proteins of *Bacillus* sp. (Table 2) with unknown function. As no alternative ORFs are present that could explain the observed activity, we conclude that the corresponding proteins, which are too small to constitute (complete) transporter proteins, complement the deficiency of the host strain, either as novel transcription activators (false positives) or as members of new classes of secreted glycosidases (true positives). ORF1, located on pSEM2, carries a signal sequence predicted to lead to protein export to the periplasm, which is an indication that this ORF encodes a secreted glycosidase or hydrolase. Furthermore, pSEM2 can convert X-Gal, which is known to reach the cytoplasm in a non-phosphorylated form. Since both BglA and BglB can only convert phosphorylated substrates, this suggests that a heterologous enzyme is expressed, leading to the hydrolysis of the colorimetric compound. Also pSEM3 seems to encode a hydrolytic enzyme rather than a regulator protein, since ORF2 is distantly related to a group of deacylases (31 % of amino acids identical in a stretch of 97 of 394 residues in total), although no signal sequence for secretion could be detected. Unfortunately, subcloning of the three promising ORFs of pLOAM1, pSEM2, and pSEM3 in a number of different expression vectors did not yield glucovanillin-metabolizing clones so far, which may be due to expression problems or to the selection of inadequate gene starts for the preparation of the constructs. As a consequence, the proposed nature of the encoded enzymes still needs experimental verification.

3.4. Improving the selection method

Although it initially appeared to be a very attractive method due to its ease of use, the absence of background growth, and the identity of screening/selection substrate and target compound in the envisaged biocatalytic application, the selection method for glucovanillin glucosidases used in this study has obvious limitations. False-positive clones with respect to the target group of enzymes are recovered, while the isolation of true positives is limited to two situations: the cloning of a secreted glucosidase and the

(rather unlikely) combined cloning of a non-phosphorylating transporter and a glucosidase.

These problems may be overcome by using an engineered *E. coli* host, lacking the cryptic *bgl* operon and *bglA* but possessing a β-glucoside transporter. Unfortunately, the use of a strain constitutively expressing *bglS* is no solution to this problem, as transport via the PTS system inevitably results in phosphorylation of the substrate, making it unsuitable for the targeted β-glucosidases. Engineering of the *E. coli* LacY permease, in contrast, may constitute a more promising strategy. The permease transports β-galactosides into the cytoplasm driven by proton motive force and no phosphorylation occurs. Its broad specificity allows the transport of a variety of chromogenic substrates (*e.g.* X-Gal and *o*-nitrophenyl-β-D-galactoside), and by mutagenesis, LacY could be adapted to convert also maltose, i.e. accept glucose as a sugar moiety (Shuman and Beckwith, 1979). Therefore, it seems to be feasible to create an *E. coli* strain expressing a mutant permease that facilitates the passage of glucovanillin through the cell membrane, thereby allowing the effective selection of clones expressing a heterologous β-glucosidase.

4. CONCLUSIONS

Obviously, a number of questions concerning the recovered new putative glucosidases and the developed selection method itself remain open. This study, however, constitutes a lead for further investigation and revealed some of the pitfalls of growth selection assays. Besides possible restrictions related to the transport of substrates into the cell, the presence of cryptic or temporally repressed genes and operons needs to be critically evaluated when designing screening or selection methods.

8

Summary and conclusions

The discovery, characterization, and optimization of new enzymes for an ever-increasing number of conversion types and reaction conditions is a major challenge in the field of biocatalysis. With more than 2×10^6 different bacterial taxa that are expected to occur in the sea and even 4×10^6 that may be present in 1 ton of soil (Curtis et al., 2002), the microbial genes present in nature constitute a gigantic resource of potential biocatalysts – provided that one possesses the means to access them. While traditional enzyme discovery strategies that are based on the screening of pure microbial cultures can only explore a limited fraction of the microbial diversity, cultivation-independent approaches in principle provide access to the complete bacterial gene pool. To date, functional screening of gene libraries prepared by random cloning of environmental DNA is one of the most promising approaches for biocatalyst discovery that do not rely on cultivation. Since target clones are detected by their activity and no sequence information is required, this strategy may allow the discovery of completely new enzymes that use so far unknown reaction mechanisms or even catalyze novel reaction types.

In this thesis, we tackled a number of critical steps in the preparation of metagenomic expression gene banks and optimized procedures to recover new biocatalysts of two different groups of enzymes: (penicillin) amidases and glucovanillin glucosidases (Fig. 1). The excellent performance of one of the obtained amidase clones in the kinetically controlled synthesis of β-lactam antibiotics prompted us to characterize the respective enzyme (PAS2) in more detail. PAS2 was found to be a new penicillin acylase, sharing 51.4 % of its amino acid residues with the well-studied *E. coli* penicillin acylase. Due to this similarity, we were able to select a number of amino acid residues likely to be important in catalysis as targets for simultaneous site-saturation mutagenesis. With this approach, a number of improved mutants that allowed up to 5 times higher antibiotic yields than the wild-type enzyme were obtained by only screening a limited amount of mutant clones.

1. Environmental samples and DNA extraction

The basis for a successful metagenome screening project is the selection of a suitable environmental sample. In some cases, the chance of recovering target activities can be increased by the clever selection of sampling sites, at which interesting organisms are already naturally enriched. An example for this strategy is the use of vanilla pods as a starting material in the search for glucovanillin glucosidases (*Chapter 7*). We chose this sample, since the bacteria that are present on the surface of mature vanilla pods are anticipated to possess β-glucosidases that cause the release of vanillin from glucovanillin during vanilla curing. In other cases, the natural function of the target group of enzymes is obscure and, consequently, the choice of a specific environmental material is difficult, such as for penicillin acylases. In this situation, one can follow two different approaches: (1) the laboratory enrichment of organisms that can carry out the desired reaction and construction of a gene library of their collective

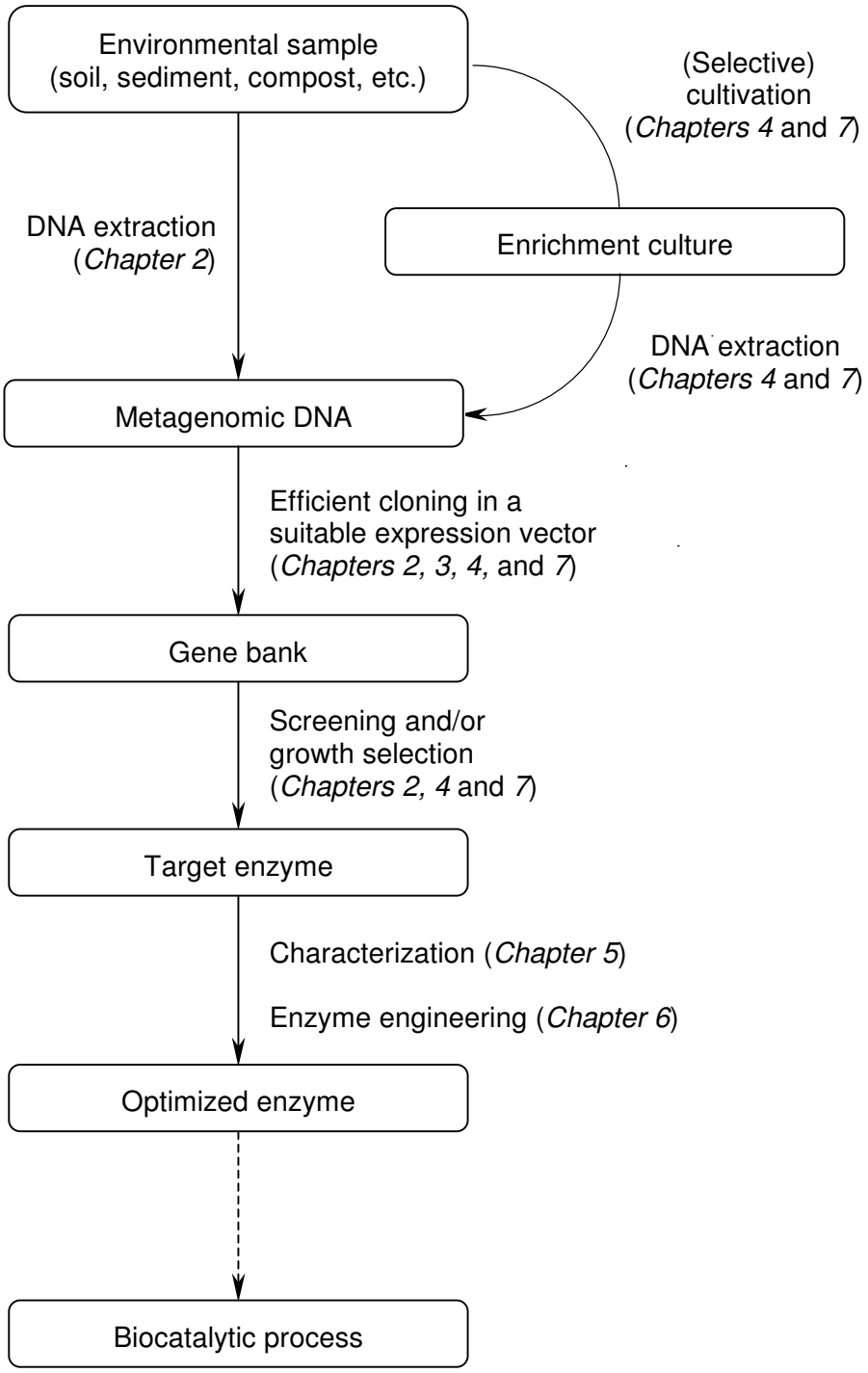


Figure 1. Providing optimized biocatalysts on basis of functional metagenome screening

DNA, or (2) the preparation of an extensive gene bank from DNA directly isolated from an environmental sample. In general, gene banks of the latter type have the advantage that they comprise a much larger bacterial diversity than enrichment gene banks and are not biased towards organisms with a certain phenotype. This implies that they can be rescreened for various different enzymatic activities. A prerequisite is the use of sample materials that accommodate a highly diverse microbial community, such as soil and sediment, and the use of efficient and non-selective DNA extraction methods. Several protocols have been developed during the past two decades for the isolation of microbial DNA from soil and sediments for phylogenetic analysis with techniques like PCR-amplification and Northern hybridization. These analytic methods, however, have very different requirements concerning e.g. DNA yield, molecular weight, or eukaryotic DNA content than procedures used for the construction of metagenomic gene libraries.

In *Chapter 2*, we thoroughly compared some of the most abundantly used DNA extraction protocols, falling into the two principal categories of environmental DNA recovery: direct and indirect DNA extraction methods. The first approach is based on the lysis of organisms directly within the sample matrix and subsequent purification of the released nucleic acids from the sample material and cell debris as originally described by Ogram et al. (1987). Although direct lysis protocols allow the recovery of large amounts of DNA and are relatively easy to use, which is why they were applied in most of the metagenome cloning projects described so far (e.g. Henne et al., 1999 and 2000; McNeil et al., 2001; Majerník et al., 2001; Rondon et al., 2000; Wang et al., 2000), we showed that the isolated DNA is less suited for the construction of metagenomic libraries. This is mainly due to the co-extraction of large portions of eukaryotic DNA (up to 93 %) that were present in the DNA extracts from a variety of sample materials. We reasoned that high amounts of co-extracted eukaryotic nucleic acids are detrimental for the quality of resulting gene banks, as they drastically increase the number of clones required for a comprehensive screening of the comprised prokaryotic activities.

In contrast to direct DNA isolation procedures, indirect protocols (pioneered by Holben et al., 1988) are based on the extraction of microbial cells from the sample matrix prior to lysis and DNA recovery. Prokaryotic cells are selectively recovered e.g. by differential centrifugation, and eukaryotic cells as well as large amounts of organic and inorganic contaminants are eliminated already in an early stage of the procedure. Consequently, the extracted DNA allows the construction of metagenomic gene banks of mostly prokaryotic DNA, and owing to its higher purity, between 2 and 3 times more clones can be obtained than with direct DNA extraction methods. The higher cloning efficiencies and the lower eukaryotic DNA contents compensate for the low total DNA yields that are generally obtained with indirect extraction methods, so that almost equal numbers of clones with prokaryotic inserts can be produced with both types of extraction protocols from the same amount of environmental starting material. The quality of gene banks obtained by cell extraction-based protocols is though significantly higher due to the absence of undesired eukaryotic DNA inserts.

2. Enrichment of target activities

A problematic feature of environmental gene banks is the generally low frequency of clones carrying a gene of interest. The preselection of organisms having a desired phenotype by classical enrichment cultivation therefore constitutes a work-saving alternative to the cloning of DNA directly isolated from environmental samples (Entcheva et al., 2001; Healy et al., 1995; Knietsch et al., 2002 and 2003a; Rees et al., 2003). Besides for the selective cultivation of amidase-containing organisms (*Chapter 4*), we also included a short enrichment step in the search for glucovanillin glucosidases (*Chapter 7*) to obtain enough cell material for DNA extraction from the sparsely colonized vanilla beans. For DNA extraction from liquid enrichment cultures, we developed a fast and non-selective protocol that yielded large amounts of high-quality DNA and allowed the parallel processing of many different samples (*Chapters 4* and *7*). As shown in *Chapter 4*, it though needs to be taken into account that due to the applied growth conditions, the diversity of the bacterial community in enrichment cultures strongly decreases, which may on one hand be due to a selective pressure that is deliberately exerted on the organisms, but which might also be caused by the more general cultivation conditions such as temperature, water activity, nutrients, etc. In this way, organisms with potentially interesting activities might be lost. In the screening for amidases, for example, 2 target clones were found in a loam soil gene bank, while no positive was recovered after enrichment.

A general advantage of a short cultivation step is the fact that eukaryotic organisms can either not grow at all in the employed growth media or have long proliferation times, which eliminates the need to develop protocols that are selective for prokaryotic DNA and allows the construction of comparatively small gene banks. In order to access a larger number of different types of organisms, several small enrichment gene banks can easily be constructed in parallel, which we consider to be a very efficient way of access to novel enzymatic activities (*Chapter 4*). Even when gene banks may be skewed towards easily cultivable bacterial species, the risk of redundancy in gene recovery seems to be negligible in view of the immense and almost unexplored pool of microbial genes present in nature.

Although growth selection in liquid culture turned out to be a useful tool for the enrichment of desired activities in a gene bank, it should be noted that solely the presence of a target gene does not yet confer the ability to grow on a certain substrate to its carrier organism. The gene needs to be expressed under the applied growth conditions, the encoded enzyme must be part of a complete catabolic pathway, and the organism must be competitive with other organisms present in the culture, which is influenced by a lot of different factors. Recently, an elegant approach has been proposed that aims at the selective recovery of nucleic acids encoding genes of interest directly from environmental samples without the need of a cultivation step (Schloss and Handelsman, 2003). The method is based on stable-isotope probing (SIP; Radajewski et al., 2000) and involves the direct spiking of habitats with C^{13}-labeled test substrates. Metabolically active organisms that can convert the supplied substrates incorporate the labeled carbon into their genomic DNA, which can then be separated

from non-labeled nucleic acids by ultracentrifugation due to its higher mass. Gene libraries can be prepared from the labeled DNA, i.e. solely from the fraction of organisms that are able to degrade the test substrates, which could rise the frequency of positive clones compared to gene libraries prepared from the complete metagenome.

3. Random cloning and gene expression

Whether cloning the community DNA of enrichment cultures or using metagenomic DNA directly isolated from an environmental sample, there is one crucial question: which cloning strategy results in the recovery of the broadest range of desired enzymatic activities? With functional screening methods, detection of positive clones depends on the level of expression of their heterologous target genes. Here, the minimally required expression level depends on the available screening or selection assay, i.e. the more sensitive this assay is, the lower expression may be.

We estimated the portion of bacterial genes that can in principle be recovered by random expression cloning in *E. coli* due to the presence of expression signals that are functional in this expression host (*Chapter 3*). To this end, we searched various genome sequences of taxonomically distinct bacterial species for sequence stretches that resemble *E. coli* promoters and ribosome binding sites, using a computer program termed GeneClassifier. About 40 % of the analyzed genes were found to be preceded by signal sequences that lead to expression in *E. coli*, with strong deviations between different groups of organisms. The majority of coding sequences, however, was devoid of a promoter, a ribosome binding site, or both. While genes that do not possess a ribosome binding site are extremely unlikely to be expressed even when using an expression vector, the absence of only the promoter is more likely to be compensated by the formation of a transcriptional fusion with a promoter located on the vector. In this case, still more clones must be screened to find a positive recombinant on average than when searching for a gene that is independently expressed, since the respective gene needs to be cloned behind the promoter in the correct orientation. Transcriptional fusions were found to be required for the expression of about 30 % of all genes. The number of clones that statistically needs to be screened to recover a gene of a certain expression mode can be calculated with the set of formulas derived in *Chapter 3*.

Of course, the presented estimates of genes that can be expressed independently or as transcriptional fusions are quite optimistic in view of the fact that the formation of an active enzyme is determined by more factors than the sheer presence of expression signals. Problems related to folding, secretion, posttranslational processing or the absence of required cofactors may prevent enzyme activity. To date, statistically relevant information on these aspects is too scarce to allow valid predictions on their importance in heterologous expression hosts. The presented numbers should therefore be regarded rather as upper limits than as exact numbers.

A practical implication of the *in silico* expression analysis described in *Chapter 3* concerns the insert size that should ideally be used for cloning. Since the expression of many genes will rely on the promoter located on the cloning vector, insert sizes should be as small as possible to decrease the risk that constructs carry a transcription

terminator between the vector promoter and the start of the gene. A transcription terminator would stop transcription from the vector promoter before reaching the gene of interest, and the respective clone could not be detected by functional screening. The lower limit of insert sizes that can be used is determined by the average size of the coding sequences that are targeted. Small inserts, such as typically used for cloning in high-copy vectors, are obviously no option when screening for e.g. antimicrobial or multi-step metabolic activities, which are encoded by large operons. When searching single genes or small operons, however, inserts smaller than 15 kb seem to be optimal since the resulting gene banks display the highest frequency of expressed genes per clone as revealed by our analysis.

Small insert libraries are also advantageous for the sake of their relatively easy preparation. Protocols used for the cloning and transformation of environmental DNA must be particularly efficient, given the low purity of DNA extracted from samples such as soil, sediment, or compost. We therefore developed a fast and robust procedure for the construction of small-insert metagenomic gene banks of DNA from enrichment cultures as well as from various environmental materials (Fig. 2).

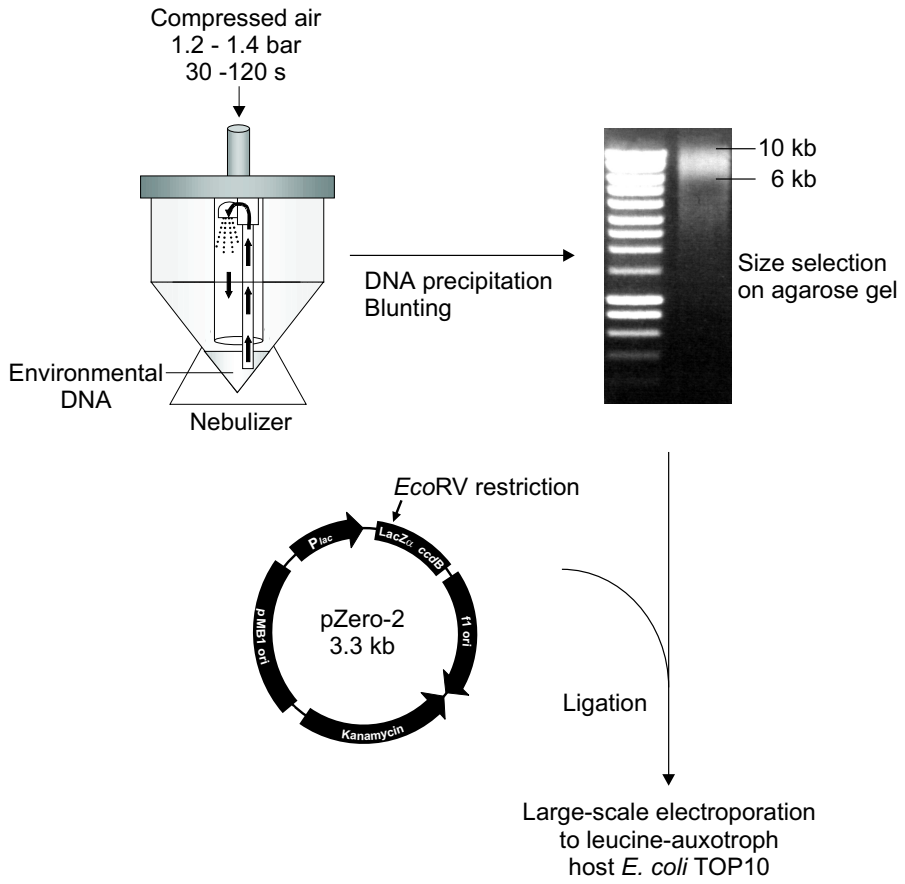


Figure 2. Optimized protocol for environmental gene bank construction used in this thesis. In a first step, environmental DNA is sheared in a nebulizer (Invitrogen) to create fragments of 6 to 10 kb. After blunting with Klenow and T4 polymerase, cured fragments are ligated in the high-copy vector pZero-2 and transformed to the *E. coli* host by electroporation. A detailed description of the protocol is given in *Chapter 4*.

DNA fragments were not prepared with an enzymatic method, since enzymatic restriction is easily inhibited by contaminants present in the DNA extracts or by incompatible methylation patterns. Instead, we fragmented the environmental DNA by mechanical shearing. After blunting and size-selection on agarose gel, fragments were ligated into a high-copy vector that allows the selection of constructs carrying an insert (pZero-2). The use of this type of vector has the advantage that background levels of clones carrying self-ligated vector molecules are kept low (≤ 5 %). Large amounts of the ligation mixture (up to 1 µg) were then used to transform *E. coli* TOP10 cells by electroporation, which was found to be the most efficient method for this host strain, typically yielding between 25,000 and 100,000 cfu per shot.

Given the stringent conditions that must be met to achieve gene expression in a heterologous host, it becomes clear that expression is one of the most important features that needs to be improved in order to increase the efficiency of metagenome mining. In view of the large number of genes that lack a promoter suitable for expression in *E. coli*, cloning vectors having strong promoters on both sides of the multiple cloning site seem to be useful, since the number of inactive clones due to wrong orientation of the insert could be halved in this way. Also cloning of metagenomic mRNA, i.e. its corresponding cDNA, behind a functional promoter appears to be an attractive strategy. Besides ensuring expression as transcriptional fusions, the possibility of creating constructs that carry a transcription terminator between the vector-promoter and the start codons of target sequences would then be ruled out. Gene banks prepared in this way will only comprise genes expressed by the studied microbial community, which may be of particular interest when screening for enzymes that allow the degradation of xenobiotics, e.g. dehalogenases or epoxide hydrolases. In this case, samples originating from contaminated sites that probably contain organisms that can metabolize the respective pollutants could be used for mRNA extraction, leading to gene banks with increased frequencies of desired genes. Unfortunately, the extraction of high-quality mRNA from environmental samples is an even more toilsome task than the isolation of DNA, and considerable efforts will need to be made to develop extraction methods that yield metagenomic mRNA suitable for reverse transcription and cloning.

4. Recovery of interesting clones from gene banks

To recover new enzymes belonging to one of our two groups of target enzymes, we used growth selection methods. In principle, these methods minimize the experimental effort needed to pick out the few positive clones present in a large gene bank, provided that the assays are properly designed. Unfortunately, the metabolic network of the host organism defines a rather rigid framework, to which potential selection protocols must be adapted. For many enzymatic activities of biotechnological interest, this is not easy. However, growth assays do not need to be absolutely selective or directly related to the ultimately desired enzyme property, as long as clones expressing a target enzyme are also recovered. In such a situation, the number of clones to be tested with a specific screening assay can be largely reduced by including

a preselection step. This strategy was successfully used in *Chapter 4* for the isolation of a novel penicillin acylase. In a first step, clones exhibiting amidase activity were preselected on basis of a hydrolytic test reaction, resulting in the recovery of 6 different positive clones. For the actually desired property, i.e. improved performance in the synthesis of β-lactam antibiotics, for which no direct growth selection method is available, only this very limited number of clones needed to be tested by HPLC to identify the active target enzyme, PAS2.

The selection of glucovanillin glucosidases, in contrast, turned out to be more difficult (*Chapter 7*). Although 9 clones able to grow on glucovanillin as the only carbon source were recovered, it appeared that the substrate might not be readily transported into the host cell. This led to the recovery of a comparatively large number of false-positive clones and compromised the direct selection of true-positives (≤ 3). These findings illustrate how the complex and often not yet well-understood metabolic activities of host organisms can interfere with selection methods, and may serve as an incentive for the critical analysis of new growth-based assay systems.

While growth selection constitutes the most efficient and elegant way of functionally assaying large gene libraries, this approach is limited to a rather narrow range of target activities. As a consequence, screening assays are widely employed to identify positive clones, although they require the separate testing of each single clone of a library. Testing of bacterial colonies on agar plates supplied with indicator reagents or test substrates that result in a color change of positive colonies is an attractive option for the screening of small- to medium-sized gene banks, although also here the number of different assays that is suitable for this format is limited. In this respect, the testing of liquid cultures in microtiter plates is the most versatile and, therefore, the most pursued screening approach. Initially, 96-well microtiter plates in combination with liquid-dispensing and plate-handling robots were used to carry out enzyme assays, mostly relying on colorimetric or fluorimetric detection of reaction products. Nowadays, 384-well plates are routinely used in many laboratories, and the use of up to 9,600-well (0.2 microliter) assay systems has been demonstrated (Oldenburg et al., 1998). In these formats, the screening of more than 10^6 clones per day is feasible. However, at this level of miniaturization, the technical limits concerning liquid handling, control of evaporation, and sensitivity of the used assay seem to be reached (Sundberg, 2000).

Functional screening is the most versatile tool for recovering positive clones from a metagenomic gene library, but also methods for the sequence-based detection of target clones are constantly being improved or newly developed. A promising sequence-based selection method, termed biopanning, was recently described and used for the isolation of dehalogenases from gene banks prepared from DNA directly isolated from soil (Gray et al., 2003). Degenerate biotinylated gene probes were hybridized to a mixture of all DNA constructs comprised in a gene library. Constructs that contained a target gene were subsequently recovered by the addition of magnetic beads that were coated with streptavidin, to which the biotinylated probes and the hybridized target constructs can selectively bind. After washing, the constructs were eluted from the magnetic beads and transformed to the desired host organism. Iterative cycles of this process were used to further enrich clones carrying a dehalogenase target

sequence. To detect active clones, the enriched libraries of limited size could then be screened with conventional function-based assays.

It is obvious that many different traits are currently being followed in order to evolve environmental expression cloning to a mature and abundantly used technology. Regarding the pace of these developments, it can be expected that a whole wealth of new enzymatic activities is on the verge of being harvested from the metagenome.

5. Enzyme improvement by semi-random mutagenesis

In *Chapter 6*, an example of the fruitful combination of metagenome screening with the improvement of a recovered enzyme by mutagenesis was presented. From an enrichment culture inoculated with sand soil, a new penicillin acylase, PAS2, was recovered (*Chapter 4*), and extensively studied with respect to its kinetic behavior (*Chapter 5*). The new enzyme and the well-studied penicillin acylase of *E. coli* have 51.4 % protein sequence identity. Compared to the *E. coli* enzyme, PAS2 was found to give significantly enhanced yields in the kinetically controlled synthesis of 6-aminopenicillanic (6-APA)-derived β-lactam antibiotics and to reach higher conversion rates. However, in the production of ampicillin, PAS2 was only advantageous at low substrate concentrations, whereas at high concentrations the *E. coli* penicillin acylase performed better. In order to improve the performance of PAS2 also in the high substrate regime, we used a semi-random mutagenesis technique. On basis of sequence comparison with the penicillin acylase of *E. coli* and a 3-D model that we constructed of the PAS2 enzyme, three amino acid residues close to the active site of PAS2 were selected as promising targets for site-directed mutagenesis. By introducing random mutations at these different sites simultaneously, a gene library of high diversity was created, focusing on parts of the protein that are expected to have a significant impact on the catalytic performance of the enzyme. As a consequence, the frequency of clearly improved mutants was high compared to that expected in a library prepared by completely random mutagenesis, with 6 improved mutants recovered from only 700 active clones that were screened. The three mutants with the highest rates in ampicillin formation were kinetically characterized and found to yield significantly more antibiotic in the high substrate concentration range than the wild-type and also than the *E. coli* enzyme. Although the activity of the three mutants in antibiotic synthesis was reduced to ≤ 20 % of the wild-type conversion rates, their high antibiotic yields as well as the low amounts of hydrolytic side-product formation make them interesting biocatalysts for the production of 6-APA derived semi-synthetic β-lactam antibiotics. Our results show that structural and mechanistic insight into the action of an enzyme is not only useful for structure-based design of improved proteins, but also for the more efficient use of directed evolution. This approach appears to be especially useful if no high-throughput screening protocols for the desired activity are available.

References

Alkema, W.B.L., de Vries, E.J., Floris, R., and D.B. Janssen (2003) Kinetics of acylation and deacylation reactions in penicillin acylase-catalyzed synthesis of β-lactam antibiotics. *Eur J Biochem* **270:** 3675-3683.

Alkema, W.B.L. (2002a) Kinetic and structural studies of penicillin acylase from *Escherichia coli*. Thesis (ISBN: 90-367-1564-4). Groningen (The Netherlands): University of Groningen.

Alkema, W.B.L., Dijkhuis, A.-J., de Vries, E.J., and D.B. Janssen (2002b) The role of hydrophobic active-site residues in substrate specificity and acyl transfer activity of penicillin acylase. *Eur J Biochem* **269:** 2093-2100.

Alkema, W.B.L., Prins, A.K., de Vries, E.J., and D.B. Janssen (2002c) Role of αArg145 and βArg263 in the active site of penicillin acylase of *Escherichia coli*. *Biochem J* **365:** 303-309.

Alkema, W.B.L., Hensgens, C.M.H., Kroezinga, E.H., de Vries, E.J., Floris, R., van der Laan, J.-M., Dijkstra, B.W., and D.B. Janssen (2000) Characterization of the β-lactam binding site of penicillin acylase of *Escherichia coli* by structural and site-directed mutagenesis studies. *Prot Eng* **13:** 857-863.

Alkema, W.B.L., Floris, R., and D.B. Janssen (1999) The use of chromogenic reference substrates for the kinetic analysis of penicillin acylases. *Anal Biochem* **275:** 47-53.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and D.J. Lipman (1990) Basic local alignment search tool. *J Mol Biol* **215:** 403-410.

Alvaro, G., Fernandez-Lafuente, R., Blanco, R.M., and J.M. Guisan (1990) Immobilization-stabilization of penicillin G acylase from *Escherichia coli. Appl Biochem Biotechnol* **26:** 181-195.

Amann, R.I. and M. Kühl (1998) *In situ* methods for assessment of microorganisms and their activities. *Curr Opin Microbiol* **1:** 352-358.

Amann, R.I., Ludwig, W., and K.-H. Schleifer (1995) Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol Rev* **59:** 143-169.

Amann, R.I., Binder, B.J., Olson, R.J., Chisholm, S.W., Devereux, R., and D.A. Stahl (1990) Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl Environ Microbiol* **56:** 1919-1925.

Arana, F. (1943) Action of a β-glucosidase in the curing of vanilla. *Food Res* **8:** 343-351.

Armstrong, D.W. and H. Yamazaki (1986) Natural flavor production: a biotechnological approach. *Trends Biotechnol* **4:** 264-268.

Arroyo, M., de la Mata, I., Acebal, C., and M.P. Castillón (2003) Biotechnological applications of penicillin acylases: state-of-the-art. *Appl Microbiol Biotechnol* **60:** 507-514.

Barbero, J.L., Buesa, J.M., Gonzalez de Buitrago, G., Mendez, E., Pez-Aranda, A., and J.L. Garcia (1986) Complete nucleotide sequence of the penicillin acylase from *Kluyvera citrophila*. *Gene* **49:** 69-80.

Bassam, J.B., Caetano-Anollés, G., and P.M. Gresshof (1991) Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal Biochem* **196:** 80-83.

Béjà, O., Suzuki, M.T., Heidelberg, J.F., Nelson, W.C., Preston, C.M., Hamada, T., Eisen, J.A., Fraser, C.M., and E.F. DeLong (2002) Unsuspected diversity among marine aerobic anoxygenic phototrophs. *Nature* **415:** 630-633.

Béjà, O., Suzuki, M.T., Koonin, E.V., Aravind, L., Hadd, A., Nguyen, L.P., Villacorta, R., Amjadi, M., Garrigues, C., Jovanovich, S.B., Feldman, R.A., and E.F. DeLong (2000) Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environ Microbiol* **2:** 516-529.

Bintrim, S.B., Donohue, T.J., Handelsman, J., Roberts, G.P., and R.M. Goodman (1997) Molecular phylogeny of Archaea from soil. *Proc Natl Acad Sci* **94:** 277-282.

Bloem, J., Veninga, M., and J. Shepherd (1994) Fully automatic determination of soil bacterial numbers, cell volumes, and frequencies of dividing cells by confocal laser scanning microscopy and image analysis. *Appl Environ Microbiol* **61:** 926-936.

Bodenteich, A., Chissoe, S., Wang, Y.-F., and B.A. Roe (1994) Automated DNA sequencing and analysis. San Diego (USA): Academic Press.

Bohannan, B.J.M. and J. Hughes (2003) New approaches to analyzing microbial biodiversity data. *Curr Opin Microbiol* **6:** 282-287.

Bollag, J.-M., Shuttleworth, K.L., and D.H. Anderson (1988) Laccase-mediated detoxification of phenolic compounds. *Appl Environ Microbiol* **54:** 3086-3091.

Borneman, J., Skroch, P.W., O'Sullivan, K.M., Palus, J.A., Rumjanek, N.G., Jansen, J.L., Nienhuis, J., and E.W. Triplett (1996) Molecular microbial diversity of an agricultural soil in Wisconsin. *Appl Environ Microbiol* **62:** 1935-1943.

Brady, S.F. and J. Clardy (2000) Long-chain *N*-acyl amino acid antibiotic isolated from heterologously expressed environmental DNA. *J Am Chem Soc* **122:** 12903-12904.

Brannigan, J.A., Dodson, G.G., Done, S.H., Hewitt, L., McVey, C.E., and K.S. Wilson (2000) Structural studies of penicillin acylase. *Appl Biochem Biotechnol* **88:** 313-319.

Brannigan J.A., Dodson, G., Duggleby, H.J., Moody, P.C.E, Smith, J.L., Tomchick, D.R., and A.G. Murzin. (1995) A protein catalytic framework with an N-terminal nucleophile is capable of self-activation. *Nature* **378:** 416-419.

Brown, J.W., Daniels, C.J., and J.N. Reeve (1989) Gene structure, organization, and expression in archaeabacteria. *CRC Crit Rev Microbiol* **16:** 287-337.

Bruggink, A. and P.D. Roy (2001) Industrial synthesis of semi-synthetic antibiotics. *In*: Synthesis of β-lactam antibiotics. Chemistry, Biocatalysis & Process Integration (Bruggink, A., ed.). Dordrecht (The Netherlands): Kluwer Academic Publishers, pp. 13-55.

Bruggink, A., Roos, E.C., and E. de Vroom (1998) Penicillin acylase in the industrial production of β-lactam antibiotics. *Org Proc Res & Develop* **2:** 128-133.

Brunerie, P.M. (1998) Process for the production of natural vanilla extract by enzymatic processing of green vanilla pods, and extract thereby obtained. US Patent (number 5,705,205).

Burton, S.G., Cowan, D.A., and J.M. Woodley (2002) The search for the ideal biocatalyst. *Nature Biotechnol* **20:** 37-45.

Buta, C., Lämmle, K., Zipper, H., Flemming, P., and M. Breuer (2003). New technical enzymes from environmental gene libraries. Darmstadt (Germany): Metagenomics Workshop.

Casamayor, E.O., Schäfer, H., Bañeras, L., Pedrós-Alió, C. and Muyzer, G. (2000) Identification of and spatio-temporal differences between microbial assemblages from two neighboring sulfurous lakes: comparison by microscopy and denaturing gradient gel electrophoresis. *Appl Environ Microbiol* **66:** 499-508.

Cerrutti, P. and S.M. Alzamora (1997) Vanillin as an antimicrobial for producing shelf-stable strawberry puree. *J Food Sci* **62:** 608-610.

Cole, M. (1969) Penicillins and other acylamino compounds synthesized by the cell-bound penicillin acylase of *Escherichia coli*. *Biochem J* **115:** 747-756.

Cottrell, M.T., Moore, J.A., and D.L. Kirchman (1999) Chitinases from uncultured marine microorganisms. *Appl Environ Microbiol* **65:** 2553-2557.

Courtois, S., Cappellano, C.M., Ball, M., Francou, F.-X., Normand, P., Helynck, G., Martinez, A., Kolvek, S.J., Hopke, J., Osburne, M.S., August, P.R., Nalin, R., Guérineau, M., Jeannin, P., Simonet, P., and J.-L. Pernodet (2003) Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Appl Environ Microbiol* **69:** 49-55.

Cowan, D.A. (2000) Microbial genomes - the untapped resource. *Trends Biotechnol* **18:** 14-16.

Crameri, A., Raillard, S.A., Bermudez, E., and W.P.C. Stemmer (1998) DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* **391:** 288-291.

Curtis, T.P., Sloan, W.T., and J.W. Scannell (2002) Estimating prokaryotic diversity and its limits. *Proc Natl Acad Sci USA* **99:** 10494-10499.

Daniel, R. (2002). Construction of environmental libraries for functional screening of enzyme activity. *In*: Directed molecular evolution of proteins (Brakmann, S. and K. Johnsson, eds.). Weinheim (Germany): Wiley-VCH, pp. 63-77.

de Lorenzo, V. and W.W. Mohn (2003). Genetic traps for surveying the biodegradation landscape of polluted sites. Darmstadt (Germany): Metagenomics Workshop.

Demain, A.L. (2000) Small bugs, big business: the economic power of the microbe. *Biotechnol Adv* **18:** 499-514.

Done, S.H., Brannigan, J.A., Moody, P.C.E., and R.E. Hubbard (1998) Ligand-induced conformational change in penicillin acylase. *J Mol Biol* **284:** 463-475.

Duarte, G.F., Rosado, A.S., Seldin, L., Keijzer-Wolters, A.C., and J.D. van Elsas (1997) Extraction of ribosomal RNA and genomic DNA from soil for studying the diversity of the indigenous bacterial community. *J Microbiol Methods* **32:** 21-29.

Duggleby, H.J., Tolley, S.P., Hill, C.P., Dodson, E.J., Dodson, G., and P.C.E. Moody (1995) Penicillin acylase has a single-amino-acid catalytic centre. *Nature* **373:** 264-268.

Edwards, A.P. and J.M. Bremner (1965) Dispersion of mineral colloids using cation-exchange resins. *Nature* **205:** 208-209.

Edwards, K., Johnstone, C., and C. Thompson (1991) A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucl Acids Res* **19:** 1349.

Elander, R.P. (2003) Industrial production of β-lactam antibiotics. *Appl Microbiol Biotechnol* **61:** 385-392.

Engelen, B., Heuer, H., Felske, A., Nübel U., Smalla, K., and H. Backhaus (1995) Protocols for TGGE. Braunschweig (Germany): Workshop on application of DGGE and TGGE in microbial ecology, Biologische Bundesanstalt für Land- und Forstwirtschaft.

Entcheva, P., Liebl, W., Johann, A., Hartsch, T., and W.R. Streit (2001) Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia. *Appl Environ Microbiol* **67:** 89-99.

Ermolaeva, M.D., Khalak, H.G., White, O., Smith, H.O., and S.L. Salzberg (2000) Prediction of transcription terminators in bacterial genomes. *J Mol Biol* **301:** 27-33.

Essack, S.Y. (2001) The development of β-lactam antibiotics in response to the evolution of β-lactamases. *Pharm Res* **18:** 1391-1399.

Fernández-Lafuente, R., Rosell, C.M., and J.M. Guisan (1996) Dynamic reaction design of enzymic biotransformations in organic media: equilibrium-controlled synthesis of antibiotics by penicillin acylase. *Biotechnol Appl Biochem* **24:** 139-143.

Firestine, S.M., Salinas, F., Nixon, A.E., Baker, S.J., and S.J. Benkovic (2000) Using an AraC-based three-hybrid system to detect biocatalysts *in vivo*. *Nature Biotechnol* **18:** 544-547.

Forney, L.J. and D.C.L. Wong (1989) Alteration of the catalytic efficiency of penicillin amidase from *Escherichia coli*. *Appl Environ Microbiol* **55:** 2556-2560.

Garmendia, J., Devos, D., Valencia, A., and V. de Lorenzo (2001) *A la carte* transcriptional regulators: unlocking responses of the prokaryotic enhancer-binding protein XylR to non-natural effectors. *Mol Microbiol* **42:** 47-59.

Gelsomino, A., Keijzer-Wolters, A.C., Cacco, G., and J.D. van Elsas (1999) Assessment of bacterial community structure in soil by polymerase chain reaction and denaturing gradient gel electrophoresis. *J Microbiol Methods* **38:** 1-15.

Gillespie, D.E., Brady, S.F., Bettermann, A.D., Cianciotto, N.P., Liles, M.R., Rondon, M.R., Clardy, J., Goodman, R.M., and J. Handelsman (2002) Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. *Appl Environ Microbiol* **68:** 4301-4306.

Gold, L., Pribnow, D., Schneider, T., Shinedling, S., Singer, B., and G. Stormo (1981) Translational initiation in prokaryotes. *Annu Rev Microbiol* **35:** 365-403.

Gray, K.A., Richardson, T.H., Robertson, D.E., Swanson, P.E., and M.V. Subramanian (2003) Soil-based gene discovery: a new technology to accelerate and broaden biocatalytic applications. *Adv Appl Microbiol* **52:** 1-27.

Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J., and R.M. Goodman (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* **5:** R245-R249.

Healy, F.G., Ray, R.M., Aldrich, H.C., Wilkie, A.C., Ingram, L.O., and K.T. Shanmugam (1995) Direct isolation of functional genes encoding cellulases from the microbial consortia in a thermophilic, anaerobic digester maintained on lignocellulose. *Appl Microbiol Biotechnol* **43:** 667-674.

Hedstrom, L., Szilagyi, L., and W.J. Rutter (1992) Converting trypsin to chemotrypsin: the role of surface loops. *Science* **255:** 1249-1253.

Helfand, M.S., Hujer, A.M., Sönnichsen, F.D., and R.A. Bonomo (2002) Unexpected advanced generation cephalosporinase activity of the M69F variant of SHV β-lactamase. *J Biol Chem* **277:** 47719-47723.

Henne, A., Schmitz, R.A., Bömeke, M., Gottschalk, G., and R. Daniel (2000) Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli*. *Appl Environ Microbiol* **66:** 3113-3116.

Henne, A., Daniel, R., Schmitz, R.A., and G. Gottschalk (1999) Construction of environmental DNA libraries in *Escherichia coli* and screening for the presence of genes conferring utilization of 4-hydroxybutyrate. *Appl Environ Microbiol* **65:** 3901-3907.

Henrik, N., Engelbrecht, J., Brunak, S., and G. von Heijne (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng* **10:** 1-6.

Hernández-Jústiz, O., Terreni, M., Pagani, G., García, J.L., Guisán, J.M., and R. Fernández-Lafuente (1999) Evaluation of different enzymes as catalysts for the production of β-lactam antibiotics following a kinetically controlled strategy. *Enzyme Microb Technol* **25:** 336-343.

Heuer, H. and K. Smalla (1997) Application of denaturing gradient gel electrophoresis and temperature gradient electrophoresis for studying soil microbial communities. *In*: Modern Soil Microbiology (van Elsas, J.D., Trevors, J.T. and E.M.H. Wellington, eds.). New York (USA): Marcel Dekker, pp. 353-373.

Holben, W.E., Jansson, J.K., Chelm, B.K., and J.M. Tiedje (1988) DNA probe method for the detection of specific microorganisms in the soil bacterial community. *Appl Environ Microbiol* **54:** 703-711.

Holmes, A.J., Bowyer, J., Holley, M.P., O'Donoghue, M., Montgomery, M., and M.R. Gillings (1999) Characterization of methanotrophic bacterial populations in soils showing atmospheric methane uptake. *Appl Environ Microbiol* **65:** 3312-3318.

Hopkins, D.W., MacNaughton, S.J., and A.G. O'Donnell (1991) A dispersion and differential centrifugation technique for representatively sampling microorganisms from soil. *Soil Biol Biochem* **23:** 217-225.

Horn, M., Collingro, A., Schmitz-Esser, S., Purkhold, U., Beier, C., Fartmann, B., Brandt, P., Nyakatura, G.J., Droege, M., Frishman, G., Montrone, C., Frishman, D., Rattei, T., Mewes, H.-W., and M. Wagner (2003). EDGE - The environmental chlamydiae genome project. Darmstadt (Germany): Metagenomics Workshop.

Huber, F.M., Chauvette, R.R., and B.G. Jackson (1972). Chemistry and biology of β-lactam antibiotics. New York (USA): Academic Press.

Hugenholtz, P., Goebel, B.M., and N.R. Pace (1998) Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J Bacteriol* **180:** 4765-4774.

Jacobsen, C.S. and O.F. Rasmussen (1992) Development and application of a new method to extract bacterial DNA from soil based on separation of bacteria from soil with cation-exchange resin. *Appl Environ Microbiol* **58:** 2458-2462.

James, M.N., Sielecki, A.R., Brayer, G.D., Delbaere, L.T., and C.A. Bauer (1980) Structures of product and inhibition complexes of *Streptomyces griseus* protease A at 1.8 Å resolution. A model for serine protease catalysis. *J Mol Biol* **144:** 43-88.

Janssen, D.B., Scheper, A., and B. Witholt (1984) Biodegradation of 2-chloroethanol by pure bacterial cultures. *Prog Ind Microbiol* **29:** 169-178.

Kaeberlein, T., Lewis, K., and S.S. Epstein (2002) Isolating "uncultivable" microorganisms in pure culture in a simulated natural environment. *Science* **296:** 1127-1131.

Kamoda, S. and Y. Saburi (1993) Structural and enzymatical comparison of lignostilbene-α,β-dioxygenase isozymes, I, II, and III, from *Pseudomonas paucimobilis* TMY1009. *Biosci Biotechnol Biochem* **57:** 931-934.

Kasche, V., Lummer, K., Nurk, A., Piotraschke, E., Rieks, A., Stoeva, S., and W. Voelter (1999) Intramolecular autoproteolysis initiates the maturation of penicillin amidase from *Escherichia coli*. *Biochim Biophys Acta* **1433:** 76-86.

Kasche, V. (1986) Mechanism and yields in enzyme catalyzed equilibrium and kinetically controlled synthesis of β-lactam antibiotics, peptides and other condensation products. *Enzyme Microb Technol* **8:** 4-16.

Kast, P., Asif-Ullah, M., Jiang, N., and D. Hilvert (1996) Exploring the active site of chorismate mutase by combinatorial mutagenesis and selection: The importance of electrostatic catalysis. *Proc Natl Acad Sci USA* **93:** 5043-5048.

Kazlauskas, R.J. (2000) Molecular modeling and biocatalysis: explanations, predictions, limitations, and opportunities. *Curr Opin Chem Biol* **4:** 81-88.

Klappenbach, J.A., Saxman, P.R., Cole, J.R., and T.M. Schmidt (2000) rrndb: the ribosomal RNA operon copy number database. *Nucleic Acids Res* **29:** 181-184.

Knietsch, A., Bowien, S., Withed, G., Gottschalk, G., and R. Daniel (2003a) Indentification and characterization of coenzyme B_{12}-dependent glycerol dehydratase- and diol dehydratase-encoding genes from metagenomic DNA libraries derived from enrichment cultures. *Appl Environ Microbiol* **69:** 3048-3060.

Knietsch, A., Waschkowitz, T., Bowien, S., Henne, A., and R. Daniel (2003b) Metagenomes of complex microbial consortia derived from different soils as sources of novel genes conferring formation of carbonyls from short-chain polyols on *Escherichia coli*. *J Mol Microbiol Biotechnol* **5:** 46-56.

Knietsch, A., Waschkowitz, T., Bowien, S., Henne, A., and R. Daniel (2002) Construction and screening of metagenomic libraries derived from enrichment cultures: generation of a gene bank for genes conferring alcohol oxidoreductase activity on *Escherichia coli*. *Appl Environ Microbiol* **69:** 1408-1416.

Kresk, M. and E.M.H. Wellington (1999) Comparison of different methods for the isolation and purification of total community DNA from soil. *J Microbiol Methods* **39:** 1-16.

Krings, U. and R.G. Berger (1998) Biotechnological production of flavors and fragrances. *Appl Microbiol Biotechnol* **49:** 1-8.

Kuske, C.R., Banton, K.L., Adorada, D.L., Stark, P.C., Hill, K.K., and P.L. Jackson (1998) Small-scale DNA sample preparation method for field PCR detection of microbial cells and spores in soil. *Appl Environ Microbiol* **64:** 2463-2472.

Kutzbach, C. and E. Rauenbusch (1974) Preparation and general properties of crystalline penicillin acylase from *Escherichia coli* ATCC 11105. *Hoppe-Seyler's Z Physiol Chem* **354:** 45-53.

Lebbink, J.H.G., Kaper, T., Bron, P., van der Oost, J., and W.M. de Vos (2000) Improving low-temperature catalysis in the hyperthermostable *Pyrococcus furiosus* β-glucosidase CelB by directed evolution. *Biochemistry* **39:** 3656-3665.

Leemhuis, H., Rozeboom, H.J., Wilbrink, M., Euverink, G.J., Dijkstra, B.W., and L. Dijkhuizen (2003) Conversion of cyclodextrin glycosyltransferase into a starch hydrolase by directed evolution: the role of alanine 230 in acceptor subsite +1. *Biochemistry* **42:** 7518-7526.

Leff, L.G., Dana, J.R., McArthur, J.V., and L.J. Shimkets (1995) Comparison of methods of DNA extraction from stream sediments. *Appl Environ Microbiol* **61:** 1141-1143.

Lenhard, B. and W.W. Wasserman (2002) TFBS: Computational framework for transcription factor binding site analysis. *Bioinformatics* **18:** 1135-1136.

Lesage-Meessen, L., Delattre, M., Haon, M., Thibault, J.-F., Ceccaldi, B.C., Brunerie, P., and M. Asther (1996) A two-step bioconversion process for vanillin production from ferulic acid combining *Aspergillus niger* and *Pycnoporus cinnabarinus. J Biotechnol* **50:** 107-113.

Lindahl, V. and L.R. Bakken (1995) Evaluation of methods for extraction of bacteria from soil. *FEMS Microbiol Ecol* **16:** 135-142.

Lingen, B., Kolter-Jung, D., Dunkelmann, P., Feldmann, R., Grotzinger, J., Pohl, M., and M. Muller (2003) Alteration of the substrate specificity of benzoylformate decarboxylase from *Pseudomonas putida* by directed evolution. *Chembiochem* **4:** 721-726.

Lisser, S. and H. Margalit (1993) Compilation of *E. coli* mRNA promoter sequences. *Nucleic Acids Res* **21:** 1507-1516.

Lloyd-Jones, G. and D.W.F. Hunter (2001) Comparison of rapid DNA extraction methods applied to contrasting New Zealand soils. *Soil Biol Biochem* **33:** 2053-2059.

Lorenz, P., Liebeton, K., Niehaus, F., and J. Eck (2002) Screening for novel enzymes for biocatalytic processes: accessing the metagenome as a resource of novel functional sequence space. *Curr Opin Biotechnol* **13:** 572-577.

Lorenz, P. and C. Schleper (2002) Metagenome - a challenging source of enzyme discovery. *J Mol Catal B:Enz* **19-20:** 13-19.

Lorenz, P., Liebeton, K., Niehaus, F., Eck, J., and H. Zinke (2001). Novel enzymes from unknown microbes - direct cloning of the metagenome. Darmstadt (Germany): 5th International Symposium on Biocatalysis and Biotransformation.

MacNeil, I.A., Tiong, C.L., Minor, C., August, P.R., Grossman, T.H., Loiacono, K.A., Lynch, B.A., Phillips, T., Narula, S., Sundaramoorthi, R., Tyler, A., Aldredge, T., Long, H., Gilman, M., Holt, D., and M.S. Osburne (2001) Expression and isolation of antimicrobial small molecules from soil DNA libraries. *J Mol Microbiol Biotechnol* **3:** 301-308.

Maidak, B.L., Cole, J.R., Parker Jr., C.T., Garrity, G.M., Larsen, N., Li, B., Lilburn, T.G., McCaughey, M.J., Olsen, G.J., Overbeek, R., Pramanik, S., Schmidt, T.M., Tiedje, J.M., and C.R. Woese (1999) A new version of the RDP (Ribosomal Database Project). *Nucleic Acids Res* **27:** 171-173.

Majerník, A., Gottschalk, G., and R. Daniel (2001) Screening of environmental DNA libraries for the presence of genes conferring $Na^+(Li^+)/H^+$ antiporter activity on *Escherichia coli*: characterization of the recovered genes and the corresponding gene products. *J Bacteriol* **183:** 6645-6653.

Maleszka, R. and G.D. Clark-Walker (1993) Yeasts have a four-fold variation in ribosomal DNA copy number. *Yeast* **9:** 53-58.

Marchesi, J.R. and A.J. Weightman (2003) Comparing the dehalogenase gene pool in cultivated α-halocarboxylic acid-degrading bacteria with the environmental metagene pool. *Appl Environ Microbiol* **69:** 4375-4382.

Miller, C.A. (2000) Advances in enzyme discovery technology: capturing diversity. *Biotechnol Reports* **11:** 489-495.

Miller, D.N., Bryant, J.E., Madsen, E.L., and W.C. Ghiorse (1999) Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. *Appl Environ Microbiol* **65:** 4715-4724.

Minagawa, H., Shimada, J., and H. Kaneko (2003) Effect of mutations at Glu160 and Val198 on the thermostability of lactate oxidase. *Eur J Biochem* **270:** 3628-3633.

Möller, E.M., Bahnweg, G., Sandermann, H., and H.H. Geiger (1992) A simple and efficient protocol for isolation of high molecular weight DNA from filamentous fungi, fruit bodies, and infected plant tissues. *Nucleic Acids Res* **20:** 6115-6116.

Muyzer, G., de Waal, E.C., and A.G. Uitterlinden (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl Environ Microbiol* **59:** 695-700.

Nakai K. and P. Horton (1999) PSORT: a program for detecting the sorting signals of proteins and predicting their subcellular localization. *Trends Biochem Sci* **24:** 34-35.

Nübel, U., Garcia-Pichel, F., Kühl, M., and G. Muyzer (1999) Quantifying microbial diversity: morphotypes, 16S rRNA genes, and carotenoids of oxygenic phototrophs in microbial mats. *Appl Environ Microbiol* **65:** 422-430.

Oefner, P.J., Hunicke-Smith, S.P., Chiang, L., Dietrich, F., Mulligan, J., and R.W. Davis (1996) Efficient random subcloning of DNA sheared in a recirculating point-sink flow system. *Nucleic Acids Res* **24:** 3879-3886.

Ogram, A., Sayler, G.S., and T. Barkay (1987) The extraction and purification of microbial DNA from sediments. *J Microbiol Methods* **7:** 57-66.

Oldenburg, K., Zhang, J., Chen, T., Maffia, A., Blom, K., Combs, A., and T. Chung (1998) Assay miniaturization for ultra-high throughput screening of combinatorial and discrete compound libraries: a 9600-well (0.2 microliter) assay system. *J Biomol Screening* **3:** 55-62.

Oliver, J.D., Nilsson, L., and S. Kjelleberg (1991) Formation of non-culturable *Vibrio vulnificus* cells and its relationship to the starvation state. *Appl Environ Microbiol* **57:** 2640-2644.

Olsen, M.J., Stephens, D., Griffiths, D., Daugherty, P., Georgiou, G., and B.L. Iverson (2000) Function-based isolation of novel enzymes from a large library. *Nat Biotechnol* **18:** 1071-1074.

Otten, L.G., Soi, C.F., Vrielink, J., Cool, R.H., and W.J. Quax (2002) Altering the substrate specificity of cephalosporin acylase by directed evolution of the β-subunit. *J Biol Chem* **277:** 42121-42127.

Page, R.D.M. (1996) An application to display phylogenetic trees on personal computers. *Comput Appl Biosci* **12:** 357-358.

Pielou, E.C. (1977) Ecological diversity and its measurement. *In*: Mathematical Ecology. New York (USA): Wiley-Interscience, pp. 292-311.

Polgar, L. (1989) Mechanism of protease action. Cambridge (UK): CRC.

Priefert, H., Rabenhorst, J., and A. Steinbüchel (2001) Biotechnological production of vanillin. *Appl Microbiol Biotechnol* **56:** 296-314.

Radajewski, S., Ineson, P., Parekh, N.R., and J.C. Murrell (2000) Stable-isotope probing as a tool in microbial ecology. *Nature* **403:** 646-649.

Radomski, C., Seow, K., Warren, R., and W. Yap (1988) Method for isolating xylanase gene sequences from soil DNA, compositions useful in such method and compositions obtained thereby. US Patent (number 5,849,491).

Ramachandra Rao, S. and G.A. Ravishankar (2000) Vanilla flavor: production by conventional and biotechnological routes. *J Sci Food Agric* **80:** 289-304.

Ranadive, A.S. (1994). Vanilla-cultivation, curing, chemistry, technology and commercial products. *In*: Developments in food science (G. Charalambous, ed.). Amsterdam (The Netherlands): Elsevier Science Publishers BV, vol. 34, pp. 517-577.

Ranadive, A.S. (1992) Vanillin and related flavor compounds in vanilla extracts made from beans of various global origins. *J Agric Food Chem* **40:** 1922-1924.

Rees, H.C., Grant, S., Jones, B., Grant, W.D., and S. Heaphy (2003) Detecting cellulase and esterase enzyme activities encoded by novel genes present in environmental DNA libraries. *Extremophiles* **4:** 415-421.

Richardson, T.H., Tan, X., Frey, G., Callen, W., Cabell, M., Lam, D., Macomber, J., Short, J.M., Robertson, D.E., and C. Miller (2002) A novel, high performance enzyme for starch liquefication. *J Biol Chem* **277:** 26501-26507.

Röling, W.F.M., Kerler, J., Braster, M., Apriyantono, A., Stam, H., and H.W. van Verseveld (2001) Microorganisms with a taste for vanilla: microbial ecology of traditional Indonesian vanilla curing. *Appl Environ Microbiol* **67:** 1995-2003.

Rolinson, G.N., Batchelor, F.R., Butterworth, D., Cameron-Wood, J., Cole, M., Eustace, G.C., Hart, M.V., Richards, M., and E.B. Chain (1960) Formation of 6-aminopenicillanic acid from penicillin by enzymatic hydrolysis. *Nature* **187:** 236-237.

Rondon, M.R., August, P.R., Bettermann, A.D., Brady, S.F., Grossman, T.H., Liles, M.R., Loiacono, K.A., Lynch, B.A., MacNeil, I.A., Minor, C., Tiong, C.L., Gilman, M., Osburne, M.S., Clardy, J., Handelsman, J., and R.M. Goodman (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl Environ Microbiol* **66:** 2541-2547.

Rondon, M.R., Goodman, R.M., and J. Handelsman (1999) The Earth's bounty: assessing and accessing soil microbial diversity. *Trends Biotechnol* **17:** 403-409.

Roose-Amsaleg, C.L., Garnier-Sillam, E., and M. Harry (2001) Extraction and purification of microbial DNA from soil and sediment samples. *Appl Soil Ecol* **18:** 47-60.

Rosazza, J.P.N., Huang, Z., Dostal, L., Volm, T., and B. Rousseau (1995) Biocatalytic transformations of ferulic acid: an abundant aromatic natural product. *J Ind Microbiol* **15:** 457-471.

Rose, M.D., Winston, F., and P. Hieter (1990) Methods in yeast genetics: a laboratory course manual. New York (USA): Cold Spring Harbor Laboratory Press.

Roszak, D.B., Grimes, D.J., and R.R. Colwell (1984) Viable but non-recoverable stage of *Salmonella enteritidis* in aquatic systems. *Can J Microbiol* **30:** 334-338.

Ruiz-Terán, F., Perez-Amador, I., and A. López-Munguia (2001) Enzymatic extraction and transformation of glucovanillin to vanillin from vanilla green pods. *J Agric Food Chem* **49:** 5207-5209.

Salmassi, T.M. and J.R. Leadbetter (2003) Analysis of genes of tetrahydrofolate-dependent metabolism from cultivated spirochaetes and the gut community of the termite *Zootermopsis angusticollis*. *Microbiology* **149:** 2529-2537.

Sambrook, J., Fritsch, E.F., and T. Maniatis (1989). Molecular cloning: a laboratory manual. New York (USA): Cold Spring Harbor Laboratory Press.

Santegoeds, C.M., Nold, S.C., and D.M. Ward (1996) Denaturing gradient gel electrophoresis used to monitor the enrichment culture of aerobic chemoorganotrophic bacteria from a hot spring cyanobacterial mat. *Appl Environ Microbiol* **62:** 3922-3928.

Schmid, A., Dordick, J.S., Hauer, B., Kiener, A., Wubbolts M., and B. Witholt (2001) Industrial biocatalysis today and tomorrow. *Nature* **409:** 258-268.

Schink, B. (2002) Synergistic interactions in the microbial world. *Antonie van Leeuwenhoek* **81:** 257-261.

Schloss, P.D. and J. Handelsman (2003) Biotechnological prospects from metagenomics. *Curr Opin Biotechnol* **14:** 303-310.

Schneider, T.D. and R.M. Stephens (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res* **18:** 6097-6100.

Seow, K.T., Meurer, G., Gerlitz, M., Wendt-Pienkowski, E., Hutchinson, C.R., and J. Davies (1997) A study of iterative type II polyketide synthases, using bacterial genes cloned from soil DNA: a means to access and use genes from uncultured microorganisms. *J Bacteriol* **179:** 7360-7368.

Shewale J.G., Deshpande, B.S., Sudhakaran, V.K., and S.S. Ambedkar (1990) Penicillin acylases: applications and potentials. *Process Biochem Int* **25:** 97-103.

Shuman, H.A. and J. Beckwith (1979) *Escherichia coli* K-12 mutants that allow transport of maltose via the β-galactoside transport system. *J Bacteriol* **137:** 365-373.

Sio, C.F., Otten, L.G., Cool, R.H., and W.J. Quax (2003) Analysis of a substrate specificity switch residue of cephalosporin acylase. *Biochem Biophys Res Comm* **312:** 755-760.

Spratt, B.G. (1994) Resistance to antibiotics mediated by target alterations. *Science* **264:** 388-393.

Stach, J.E.M., Bathe, S., Clapp, J.P., and R.G. Burns (2001) PCR-SSCP comparison of 16S rDNA sequence diversity in soil DNA obtained using different isolation and purification methods. *FEMS Microbiol Ecol* **36:** 139-151.

Staden, R. (1983) Computer methods to locate signals in nucleic acid sequences. *Nucleic Acids Res* **12:** 509-519.

Stahl, D.A. and R. Amann (1991) Development and application of nucleic acid probes in bacterial systematics. *In*: Nucleic acid techniques in bacterial systematics (Stackebrandt, E. and M. Goodfellow, eds.). Chichester (UK): John Wiley & Sons Ltd., pp. 205-248.

Stajich, J.E., Block, D., Boulez, K., Brenner, S.E., Chervitz, S.A., Dagdigian, C., Fuellen, G., Gilbert, J.G.R., Korf, I., Lapp, H., Lehväslaiho, H., Matsalla, C., Mungall, C.J., Osborne, B.I., Pocock, M.R., Schattner, P., Senger, M., Stein, L.D., Stupka, E., Wilkinson, M.D., and E. Birney (2002) The Bioperl toolkit: perl modules for the life sciences. *Genome Research* **12:** 1611-1618.

Steffan, R.J., Goksøyr, J., Bej, A.K., and R.M. Atlas (1988) Recovery of DNA from soils and sediments. *Appl Environ Microbiol* **54:** 2908-2915.

Stokes, H.W., Holmes, A.J., Nield, B.S., Holley, M.P., Nevalainen, K.M.H., Mabbutt, B.C., and M.R. Gillings (2001) Gene cassette PCR: sequence-independent recovery of entire genes from environmental DNA. *Appl Environ Microbiol* **67:** 5240-5246.

Stormo, G.D., Schneider, T.D., and L. Gold (1986) Quantitative analysis of the relationship between sequence and functional activity. *Nucleic Acids Res* **14:** 6661-6679.

Sundberg, S.A. (2000) High-throughput and ultra-high-troughput screening: solution- and cell-based approaches. *Curr Opin Biotechnol* **11:** 47-53.

Taguchi, S., Komada, S., and H. Momose (2000) The complete amino acid substitutions at position 131 that are positively involved in cold adaptation of subtilisin BPN'. *Appl Environ Microbiol* **66:** 1410-1415.

Taran, F., Gauchet, C., Mohar, B., Meunier, S., Valleix, A., Renard, P.Y., Créminon, C., Grassi, J., Wagner, A., and C. Mioskowski (2002) High-throughput screening of enantioselective catalysts by immunoassay. *Angew Chem Int Ed* **41:** 124-127.

Tebbe, C.C. and W. Vahjen (1993) Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and a yeast. *Appl Environ Microbiol* **59:** 2657-2665.

Tien, C.C., Chao, C.C., and W.L. Chao (1999) Methods for DNA extraction from various soils: a comparison. *J Appl Microbiol* **86:** 937-943.

Torsvik, V. and L. Øvreås (2002) Microbial diversity and function in soil: from genes to ecosystems. *Curr Opin Microbiol* **5:** 240-245.

Torsvik, V., Øvreås, L., and T.F. Thingstad (2002) Prokaryotic diversity - magnitude, dynamics, and controlling factors. *Science* **296:** 1064-1066.

Torsvik, V., Daae, F.L., Sandaa, R.-A., and L. Øvreås (1998) Novel techniques for analyzing microbial diversity in natural and perturbed environments. *J Biotech* **64:** 53-62.

Torsvik, V., Goksøyr, J., and F.L. Daae (1990) High diversity in DNA of soil bacteria. *Appl Environ Microbiol* **56:** 782-787.

Torsvik, V., J. Goksøyr, and F.L. Daae (1989) High diversity in DNA of soil bacteria. *Appl Environ Microbiol* **56:** 782-787.

Valle, F., Balbás, P., Merino, E., and F. Bolivar (1991) The role of penicillin amidases in nature and industry. *Trends Biochem Sci* **16:** 36-40.

Vandamme, E.J. and W. Soetaert (2002) Bioflavors and fragrances via fermentation and biocatalysis. *J Chem Technol Biotechnol* **77:** 1323-1332.

Vellai, T. and G. Vida (1999) The origin of eukaryotes: the difference between prokaryotic and eukaryotic cells. *Proc R Soc Lond* B **226:** 1571-1577.

Vergin, K.L., Urbach, E., Stein, J.L., Delong, E.F., Lanoil, B.D., and S.J. Giovannoni (1998) Screening of a fosmid library of marine environmental genomic DNA fragments reveals four clones related to members of the order *Planctomycetales*. *Appl Environ Microbiol* **64:** 3075-3078.

von Wintzingerode, F., Göbel, U.B., and E. Stackebrandt (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol Reviews* **21:** 213-229.

Wahler, D. and J.-L. Reymond (2001) Novel methods for biocatalyst screening. *Curr Opin Chem Biol* **5:** 152-158.

Wang, G.-Y.-S., Graziani, E., Waters, B., Pan, W., Li, X., McDermott, J., Meurer, G., Saxena, G., Andersen, R.J., and J. Davies (2000) Novel natural products from soil DNA libraries in a *Streptomycete* host. *Org Lett* **2:** 2401-2404.

Washio, T., Sasayama, J., and M. Tomita (1998) Analysis of complete genomes suggests that many prokaryotes do not rely on hairpin formation in transcription termination. *Nucleic Acids Res* **26:** 5456-5463.

Yeates, C., Holmes, A.J., and M. R. Gillings (2000) Novel forms of ring-hydroxylating dioxygenases are widespread in pristine and contaminated soils. *Environ Microbiol* **2:** 644-653.

Youshko, M. I., Moody, H.M., Bukhanov, A.L., Boosten, W.H.J., and V.K. Svedas (2004) Penicillin acylase-catalyzed synthesis of β-lactam antibiotics in highly condensed aqueous systems: beneficial impact of kinetic substrate supersaturation. *Biotechnol Bioeng* **85:** 323-329.

Youshko, M.I., Chilov, G.G., Shcherbakova, T.A., and V.K. Svedas (2002a) Quantitative characterization of the nucleophile reactivity in penicillin acylase-catalyzed acyl transfer reactions. *Biochim Biophys Acta* **1599:** 134-140.

Youshko, M.I., van Langen, L.M., de Vroom, E., van Rantwijk, F., Sheldon, R.A., and V.K. Svedas (2002b) Penicillin acylase-catalyzed ampicillin synthesis using a pH gradient: a new approach to optimization. *Biotechnol Bioeng* **78:** 589-593.

Youshko, M.I., van Langen, L.M., de Vroom, E., van Rantwijk, F., Sheldon, R.A., and V.K. Svedas (2001) Highly efficient synthesis of ampicillin in an "aqueous solution-precipitate" system: repetitive addition of substrates in a semicontinuous process. *Biotechnol Bioeng* **73:** 426-430.

Youshko, M.I. and V.K. Svedas (2000) Kinetics of ampicillin synthesis catalyzed by penicillin acylase from *E. coli* in homogeneous and heterogeneous systems. Quantitative characterization of nucleophile reactivity and mathematical modeling of the process. *Biochemistry (Mosc)* **65:** 1367-1375.

Zengler, K., Toledo, G., Rappe, M., Elkins, J., Mathur, E.J., Short, J.M., and M. Keller (2002) Cultivating the uncultured. *Proc Natl Acad Sci* **99:** 15681-15686.

Zhao, H., Chockalingam, K., and Z. Chen (2002) Directed evolution of enzymes and pathways for industrial biocatalysis. *Curr Opin Biotechnol* **13:** 104-110.

Zheng, D., Alm, E.W., Stahl, D.A., and L. Raskin (1996) Characterization of universal small-subunit rRNA hybridization probes for quantitative molecular microbial ecology studies. *Appl Environ Microbiol* **62:** 4504-4513.

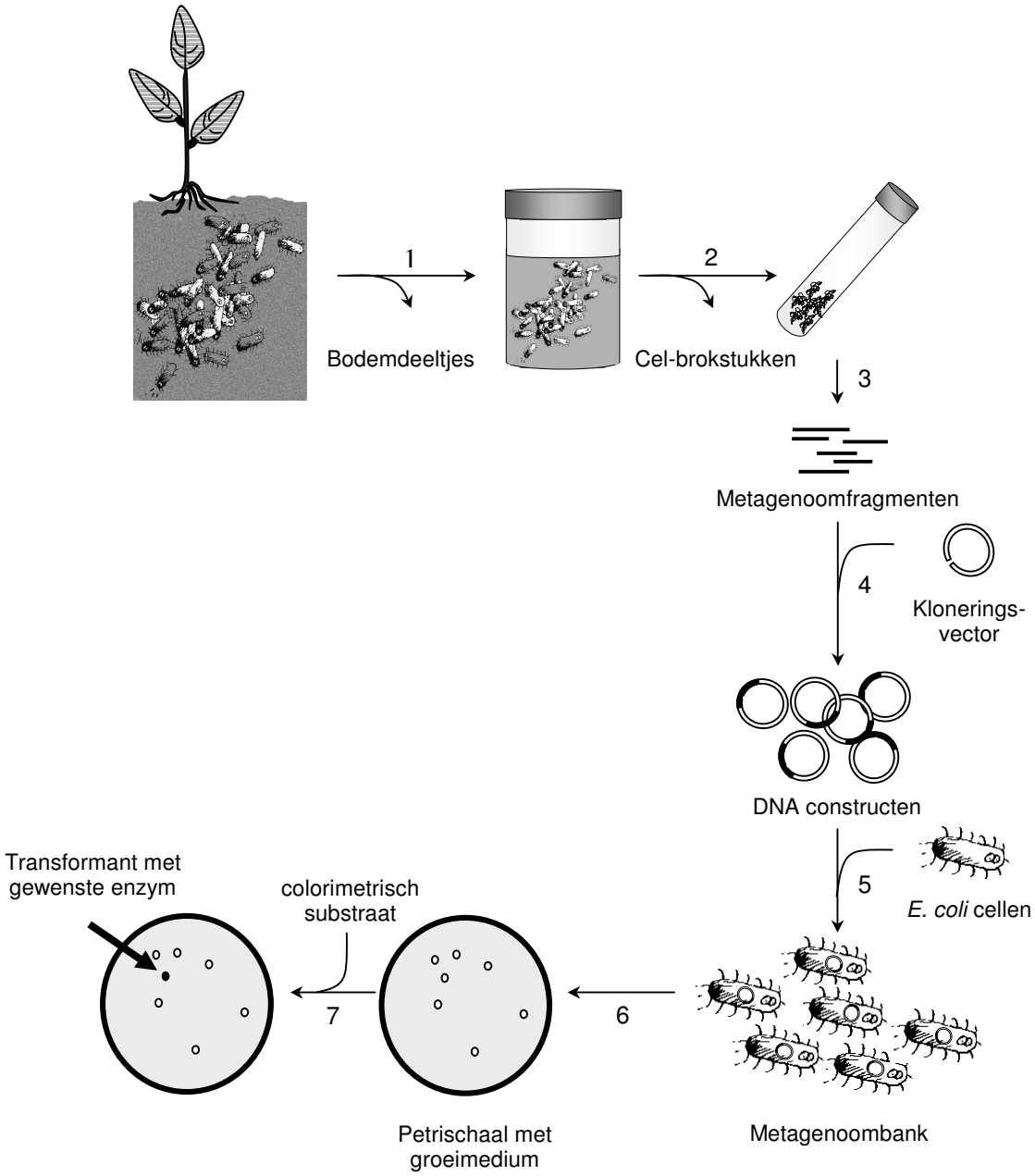
Zhou, J., Bruns, M.A., and J.M. Tiedje (1996) DNA recovery from soils of diverse composition. *Appl Env Microbiol* **62:** 316-322.

Nederlandse samenvatting

Bacteriën spelen in de biotechnologie een bijzondere rol. Naast schimmels en gisten vormen ze de basis van sommige van de oudste biotechnologische processen zoals de productie van azijn, gefermenteerde zuivelproducten, zuurkool en bij het pekelen. Niet alleen de intacte (levende) bacteriecel kan worden ingezet om bepaalde chemische reacties te doen, maar ook de uit bacteriecellen geïsoleerde afzonderlijke enzymen. Enzymen zijn katalysatoren, dat wil zeggen dat ze chemische reacties versnellen en ze behoren, op een paar uitzonderingen na, allen tot de eiwitklasse. In de industrie worden enzymen bijvoorbeeld op grote schaal toegepast in de productie van "high fructose corn sirup" (glucose isomerase), de vorming van acrylamide (nitril hydratase) en de isolatie van aminopenicillaanzuur (penicilline acylase). Het aantal van dergelijke processen stijgt snel en de behoefte aan nieuwe of verbeterde enzymen groeit navenant. De ontzagwekkende hoeveelheid van bacteriën in de natuur lijkt hierbij een schier onuitputtelijke bron van nieuwe biokatalysatoren te zijn: in slechts een gram gewone tuinaarde komen al tussen de 1.000 en 10.000 verschillende bacteriesoorten voor die uitgerust zijn met zeer diverse enzymen (dus potentiële biokatalysatoren). Echter, minder dan 1 % van alle onder de microscoop zichtbare bacteriën die in een dergelijk monster aanwezig zijn, kunnen worden gecultiveerd, hetgeen een traditionele analyse van hun enzymatische activiteit onmogelijk maakt.

Hoe kan men zich nu toch óók toegang tot de enzymen van dergelijke niet-cultiveerbare organismen verschaffen? Het antwoord ligt in de erfelijke informatie. Ook indien bepaalde organismen zich niet eenvoudig laten vermeerderen, kan men het DNA direct uit het desbetreffende monstermateriaal (bijvoorbeeld zand, tuinaarde, kompost of zuiveringsslib) isoleren en het gebruiken voor het maken van een genbank in een gemakkelijk te cultiveren bacteriële gastheerstam zoals *Escherichia coli* (Figuur 1). Omdat het gebruikte DNA niet van slechts één soort afkomstig is, maar van een complete microbiële gemeenschap, moet een navenant groot aantal transformanten (> 10^6) gemaakt worden indien men er zeker van wil zijn dat alle erfelijke informatie van die gemeenschap, het zogenaamde *metagenoom*, wordt omvat. Als de gekloneerde genen de geschikte signaalsequenties [promoters en rbs (ribosoom bindingsplaatsen)] hebben, kunnen ze in de gastheerstam tot expressie worden gebracht. Cultures of kolonies van transformanten kunnen dan door toevoegen van een substraat op de gewenste enzymactiviteit worden getest. Heeft men een recombinante *E. coli* stam met de gewenste activiteit gevonden, dan kan deze in een grotere hoeveelheid gekweekt worden en het verkregen enzym kan voor verdere karakterisatie of voor biokatalytische toepassingen geïsoleerd worden. In het kader van voorliggend proefschrift werd het screenen van metagenoombanken toegepast om nieuwe (penicilline) amidases en glucovanilline glucosidases te isoleren.

In tegenstelling tot het maken van genbanken van organismen die in reinculture kunnen worden gegroeid, is het isoleren van geschikt DNA voor een metagenoombank vaak problematisch. Met veel methodes worden slechts kleine hoeveelheden DNA van vaak slechte kwaliteit, dat wil zeggen een lage zuiverheid en laag molecuulgewicht, uit het monstermateriaal geëxtraheerd. Buiten dat – veel protocollen zijn selectief voor bepaalde groepen van bacteriën, wat resulteert in een lagere genetische diversiteit in de verkregen DNA-extracten.



Figuur 1. Het maken van een metagenoombank: Extractie van bacteriecellen uit het monstermateriaal (1) en DNA isolatie (2), fragmenteren van het metagenoom-DNA (3) en het inbrengen in de kloneringsvector (4) die als transportvehikel voor het binnensluizen van niet-eigen DNA in de *E. coli* gastheerstam dient (5). In het algemeen worden ongeveer 10^4-10^6 transformanten verkregen. De ontstane metagenoombank wordt op een geschikte vaste voedingsbodem gebracht zodat de afzonderlijke transformanten zich kunnen vermeerderen en zichtbare kolonies vormen (6). Deze kolonies kunnen dan voor elke gewenste enzymactiviteit worden gescreened, bijvoorbeeld door het toevoegen van een *colorimetrisch* testsubstraat. Dit is een substraat dat bij omzetting een verkleuring van de actieve cellen veroorzaakt. Het substraat moet zo ontwikkeld zijn dat het alleen door de transformant kan worden omgezet die het specifieke gewenste enzym tot expressie brengt (7).

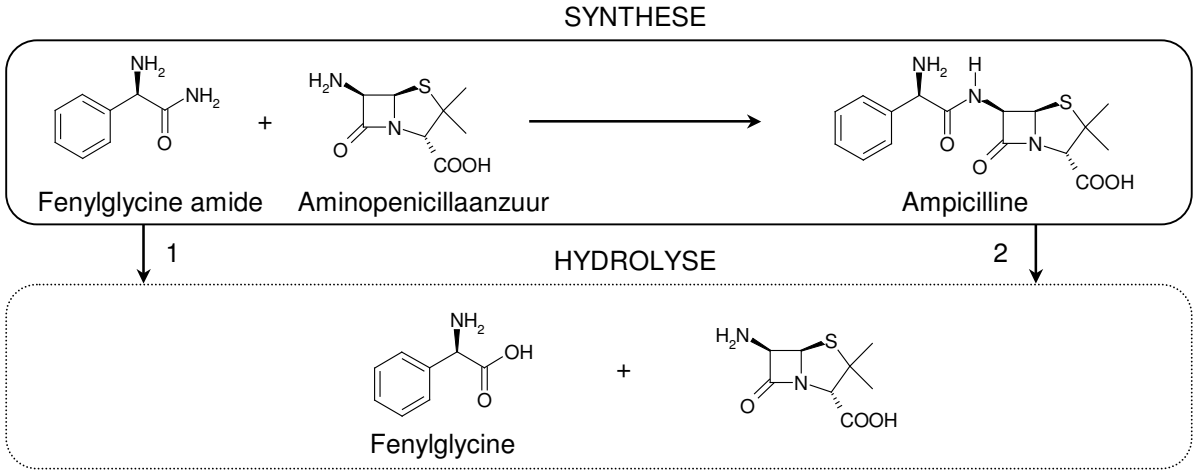
Voorts vindt er vaak een verrijking plaats met grote hoeveelheden eukaryotisch DNA dat niet in bacteriële gastheercellen tot expressie kan worden gebracht en dus tot een contaminatie van de metagenoombank met niet-productieve transformanten leidt. Om een geschikte methode voor het isoleren van metagenoom-DNA te vinden, worden in *Hoofdstuk 2* twee fundamentele manieren van DNA isolatie uit natuurlijke monsters vergeleken. Bij de zogenaamde *directe* DNA isolatie worden de aanwezige microorganismen gelyseerd terwijl ze nog in het monstermateriaal zitten en het vrijkomende DNA wordt daarna gezuiverd van groffe deeltjes en celmateriaal. Dit is in tegenstelling tot de *indirecte* DNA isolatie waarbij eerst de microbiële cellen uit het monster worden vrijgemaakt en geïsoleerd waarna in een tweede stap het DNA uit de cellen wordt verkregen (Figuur 1, stappen 1 en 2). Het bleek dat deze tweede methode DNA van beduidend betere kwaliteit gaf, wat vooral tot uiting kwam in de hogere zuiverheid en de verregaande afwezigheid van eukaryotisch DNA. Via directe weg verkregen DNA extracten bevatten daarentegen grote hoeveelheden humuszuren en tot wel 90 % eukaryotisch DNA.

Hoofdstuk 3 houdt zich bezig met de principiële vraag welke fractie van de, door het metagenoom gecodeerde enzymen, met behulp van een op activiteit gebaseerde screeningsmethode toegankelijk is. Hiervoor werden een aantal diverse bacteriële genoomsequenties met een voor dit doel geschreven computerprogramma op de aanwezigheid van expressiesignalen, die in *E. coli* werkzaam kunnen zijn, onderzocht. Deze analyse liet zien dat ongeveer 40 % van alle geanalyseerde genen een complete set van expressiesignalen bezitten en in principe in *E. coli* tot de productie van een enzym kunnen leiden. Echter, de meerderheid van alle genen mist minstens één van de twee fundamentele signaalsequenties. De afwezigheid van een promotersequentie (in ongeveer 30 % van de genen) kan daarbij gecompenseerd worden door het kiezen van een kloneringsvector met een eigen promoter. In dat geval geeft het kiezen voor kleinere DNA-fragmenten de grootste kans op een functioneel recombinant construct. Als namelijk te grote fragmenten worden gebruikt, groeit de waarschijnlijkheid dat de vorming van een mRNA transcript, die bij de vectorpromoter begint, door een voor het gen liggende transcriptieterminator wordt beëindigd zodat uiteindelijk toch geen enzym wordt gemaakt. De computeranalyse gaf te zien dat in dit opzicht een fragmentgrootte van <15 kb optimaal is. Statistisch gezien is de waarschijnlijkheid van expressie van een gen dat geen *E. coli* expressiesignalen draagt (30 % van het metagenoom) zeer laag, zelfs als zowel de promoter als de rbs van de vector worden gebruikt.

Met behulp van de in *Hoofdstukken 2* en *3* verkregen resultaten werden uiteindelijk zes metagenoombanken van verschillende uitgangsmaterialen gemaakt. Het screenen van vier van deze banken (in totaal 193.000 transformanten) op amidase-activiteit leverde zeven transformanten met de gewenste eigenschappen, zoals is beschreven in *Hoofdstuk 4*. Alle zeven kloons bevatten tot nu toe onbekende DNA-sequenties en brengen nieuwe enzymen tot expressie die met diverse substraten werden getest. Eén van de transformanten gedroeg zich bijzonder veelbelovend omdat niet alleen een breed spectrum aan amides kon worden gehydrolyseerd, maar ook omdat de synthese van penicilline G vanuit fenylazijnzuur amide en aminopenicillaanzuur kon worden gekatalyseerd. Het door deze *E. coli* stam geproduceerde enzym PAS2 – een nieuwe

vertegenwoordiger van de penicilline acylases, werd opgezuiverd en aan een gedetailleerde kinetische karakterisering onderworpen (*Hoofdstuk 5*).

Penicilline acylases kunnen, aan de ene kant, voor de hydrolyse van fermentatief geproduceerde penicilline G worden toegepast, hetgeen dient voor het maken van aminopenicillaanzuur. Aan de andere kant kunnen ze aminopenicillaanzuur ook aan een geactiveerde acyldonor koppelen en daarmee tot de synthese van semi-synthetische penicillines leiden. In deze kinetisch gecontroleerde omzetting treden twee ongewenste nevenreacties op: (1) de hydrolyse van de geactiveerde acyldonor en (2) de hydrolyse van het gevormde antibioticum (Figuur 2). De verhouding van de synthese- en hydrolysereacties, en daarmee samenhangend de maximale productopbrengst, wordt sterk bepaald door de kinetische eigenschappen van het gebruikte enzym. Op grond van hun gebrekkige synthesecapaciteiten zijn de meeste tot nog toe bekende penicilline acylases niet voor deze synthetische doeleinden geschikt. Zoals in de vergelijking met het veelbeschreven *E. coli* penicilline acylase wordt gedemonstreerd, steekt PAS2 gunstig af door zijn betere kinetische parameters met betrekking tot de synthesereactie en zijn hoge activiteit.



Figuur 2. Kinetisch gecontroleerde synthese van semi-synthetische penicillines (voorbeeld: ampicilline). Er treden twee ongewenste nevenreacties op: (1) de hydrolyse van de geactiveerde acyldonor en (2) de hydrolyse van het gevormde antibioticum.

Hoofdstuk 6 behandelt de, met behulp van moleculair biologische mutagenesemethodes, verdere optimalisering van PAS2 voor synthetische doeleinden. Op grond van de homologie van PAS2 en het *E. coli* acylase, waarvan de driedimensionale structuur bekend is, kon een ruimtelijk model van PAS2 worden geconstrueerd en aan de hand daarvan konden drie aminozuurbouwstenen worden aangewezen die vermoedelijk een directe invloed op het katalytische gedrag van het enzym hebben. Door gelijktijdige mutatie van de aminozuren op deze drie posities werd een bank van “triplemutanten” verkregen die met een geautomatiseerd HPLC-systeem op verbeterde ampicillinesynthese werd gescreened. Op deze manier konden

drie duidelijk verbeterde mutanten geïsoleerd worden die een tot vijfvoudig verhoogde productopbrengst mogelijk maakten.

In *Hoofdstuk 7* is tenslotte het screenen van 205.000 metagenoom-transformanten op glucovanilline glucosidase-activiteit beschreven. Hier werden in totaal elf positieve transformanten geïsoleerd. Voor de detectie werd een selectief groeimedium toegepast dat alleen de vermeerdering van transformanten toestaat die glucovanilline als koolstofbron kunnen gebruiken. Na evaluatie van de DNA sequenties bleek dat slechts drie van deze kloons echte glucosidases waren en dat de resterende acht transformanten genen droegen die op een andere manier in het glucovanilline-metabolisme ingrijpen. De hieruit afgeleide overwegingen geven duidelijk aan op welke veelvuldige manier het complexe metabolisme van de gastheercel met selectie- en screeningsprocedures kan interfereren.

Het vaak voorkomen van doel-activiteiten in de in dit onderzoek geteste metagenoombanken (1 positieve op 18.000 transformanten), en de nieuwheid van alle geïsoleerde DNA sequenties doet vermoeden welke grote hoeveelheid en diversiteit van nieuwe bacteriële enzymen nog op ontdekking wachten. Men mag hopen dat het aantal beschikbare, efficiënte, biokatalysatoren in de nabije toekomst snel stijgt zodat steeds nieuwe toepassingsgebieden voor biokatalytische processen kunnen worden ontsloten.

LAST BUT NOT LEAST…

I want to thank a number of people that significantly contributed to this thesis.

First of all my supervisor Dick Janssen, for the inspiring discussions, the faith in my research, and the lessons on "not writing too complex". I really learned a lot from you and enjoyed being in your lab!

I am also very grateful for the insight into the industrial point of view on biocatalysis that I got due to the members of the "STW-begeleidingscommissie", Nico Boots, Theo Sonke, Jan-Metske van der Laan, Chris Winkel, and Sven Panke. Our meetings twice per year were always very stimulating and instructive (and a nice opportunity to visit different places in The Netherlands)!

Thanks to the professors Lubbert Dijkhuizen, Wim Quax, and Dick van Elsas for the positive assessment of my thesis.

And of course, also a big smile to all the colleagues, students and guest researchers of the Janssen-group, who were always willing to share ideas, problems and beers… particularly Lixia (room mate, perfect catering during the bioinformatics course in Bochum), Bert (room mate and living chronicle, specialized in football and Wageningen history), Jin II (room mate and experienced in loempia frying), Mariël (room mate and ever-freezing predecessor at the desk close to the heating), Inez (ex-supervisor in how to get mad with fermentors), Nanne (how to survive in Beijum and ever-singing piano fan), Jolanda (organization talent and perfectionist), Wynand (ex-PA group member, now computer nerd), the biodegradation people Johan, Tjibbe, and Wouter, Simon (the last combatant on the penicillin acylase field), Martijn (almost managed to get hit by me), Erik v. H. (likes to play in the dirt and thus metagenome bank successor), Isabel (globetrotter and "Southern temperament"), Ghannia (how to create a mess in the lab…), Marco (…and how to clean it up; deserving *borrel* organizer), Maja (Not Without My Daughter), Ditte (lend a dog), Dominic (Nanne II and carnival fetishist), Chiara (best Italian coffee, but slightly failed muffins), Helen (cool haircuts), Maria (enthusiast soccer fan, even moved to Porto for the EURO 2004), Jantien (doc, post-doc, mother), Zbynek (squash and Becherovka), Peter (competitor for HPLC measuring time), Gerard (one-man research group), Jaap (still consumer of an astonishing amount of coffee), Piet (keeper of the fermentors and a great help with HPLC, robot and computer problems), Sandra E. (personified Limburgs-speaking enthusiasm), Nico (repairs just everything), Sandra H., Tamara, and Hilda (always ready to help with sending papers or filling in weird forms) and, finally, the students Freek and Hendrie, who helped to struggle with not yet optimal procedures for gene bank construction. Special thanks to Erik d. V. (sightseeing- and Baltika-guide in tropical Moscow, high proofreading activity) and Jeffrey (unforgettable camping-trip to the *rain*forest without a tent) for willing to be my paranimfen. Dear all, I will always remember the fun of the sailing week-ends, adventure trips (Six Flags and Duinrell), wine tastings, BBQs, movie evenings, *borrels*, *lab uitjes*, and the hundreds of coffee and lunch breaks we spent together!

My family and friends in “Good Old Germany”, I want to thank for their constant support, regular visits, hugh investments in telephone bills, and for reminding me that work is only a part of life. And there is of course Christian, who shared the Dutch exile (including *bitterballen*, *kaas*, and *koniginnen dag*) and the ups and downs of a PhD research project. We had a great time on our Animal Farm in the Munsterheerd, and I am sure it will become even better – wherever that may be!

And for those I forgot: don’t be offended – but get ready for the party on 9 July!

Esther

STELLINGEN

behorende bij het proefschrift van Esther M. Gabor
te verdedigen op 9 juli 2004

1. The term *uncultivable* does not refer to a well-defined intrinsic property of a bacterial organism, but to the negative outcome of an experiment, namely the attempt to enrich it.
(*Chapter 1*)

2. The nomenclature of kinetic parameters is sometimes arbitrary.
(Youshko et al., 2002, *Biochim Biophys Acta* **1599:** 134-140; *Chapters 5 and 6*)

3. *High-throughput* screening is a relative term.

4. Dit proefschrift laat zien dat het graven van een kuil in het Nederlandse strand niet alleen recreatieve doeleinden dient maar ook tot wetenschappelijke vooruitgang kan leiden. Het moge geen verbazing wekken dat de Duitse biotechnologische industrie verder is ontwikkeld dan de Nederlandse.

5. Today are *the good old times* that we'll be longing to be back in ten years.
(*Sir Peter Ustinov*)

6. The euro shows that money may create more unity between people than sublime ideals.

7. Even the smallest person can change the fortune of the world.
(*J.R.R. Tolkien, The Lord of the Rings*)

8. Insisting on tolerance leads to its extinction.

9. Ein freundliches Gesicht bringt in die Dunkelheit viel Licht.
(*Siegbert Gabor*)

Esther Gabor, juni 2004